



Universidad de
SanAndrés

Universidad de San Andrés

Maestría en Educación

Las prácticas de calificación en la educación secundaria

Un acercamiento a su validez

Autor: Cristian José Saint Germain

Mentora: Dra. Alicia W. de Camilloni

Firma de la mentora

AGRADECIMIENTOS

A mis hijos: José y Magu.

A mi familia que me mostró la maravillosa aventura de aprender y enseñar.

A María Marta, compañera y amiga en este camino.

A la Prof. Alicia Camilloni por su paciencia, confianza y sabiduría.



Universidad de
San Andrés

INDICE

AGRADECIMIENTOS.....	i
INDICE.....	ii
INDICE DE FIGURAS Y CUADROS	v
INTRODUCCIÓN.....	1
CAPÍTULO 1	
LA INVESTIGACIÓN EN EVALUACIÓN Y CALIFICACIÓN AYER Y HOY	
¿Qué aportaron más de 100 años de investigación?	
1. Los primeros estudios.....	7
2. La investigación en calificación: temas y enfoques de ayer y hoy.....	9
La validez de las calificaciones en la percepción del docente.....	10
3. Otros idiomas, ¿las mismas preocupaciones?.....	13
CAPÍTULO 2	
LA EVALUACIÓN Y LA CALIFICACIÓN EN EL PROCESO DE ENSEÑANZA Y APRENDIZAJE	
1. La evaluación: una práctica social.....	17
Evaluación educativa.....	18
2. Tipos de evaluación	
Evaluaciones externas estandarizadas.....	19
La evaluación en el aula.....	21
Evaluación formativa.....	22
Evaluación sumativa.....	23
3. Las prácticas de calificación.....	26
Las notas: su finalidad.....	31
¿Cómo se hacen, fabrican o producen? Las prácticas de calificación.....	36
Significado y validez de las calificaciones.....	46
La matriz progresiva de validez de Messick.....	49
CAPITULO 3	
HISTORIA DE LAS CALIFICACIONES	
1. ¿Para qué una historia de las calificaciones?.....	49
Un campo poco y superficialmente transitado.....	50
2. Génesis y desarrollo de la evaluación y las calificaciones.....	52
Antecedentes remotos.....	52
La bisagra: los jesuitas.....	53

La invención de la nota.....	56
3. El campo de la evaluación y el estudio de los sistemas de calificación.....	59

CAPÍTULO 4

LAS PRÁCTICAS DE CALIFICACIÓN DE LOS PROFESORES

Marco metodológico: test de escenarios de calificación de cierre de trimestre.....	63
Organización y análisis de los datos cuantitativos y sus criterios de validación.....	71
Escenarios de calificación: respuestas, comentarios y análisis	
Escenario 1.....	71
Escenario 2.....	73
Escenario 3.....	75
Escenario 4.....	77
Escenario 5.....	79
Escenario 6.....	81
Escenario 7.....	83
Escenario 8.....	84
Escenario 9.....	87
Análisis comparativo.....	88

CAPITULO 5

INTERPRETACIÓN GENERAL DE LOS ESCENARIOS DE CALIFICACIÓN

Características Generales de las prácticas de calificación.....	93
-----------------------------------------------------------------	----

CAPITULO 6

CONCLUSIONES

PRESESENTE Y FUTURO DE LAS PRÁCTICAS DE CALIFICACIÓN.....

Regímenes de evaluación y calificación en la Argentina 2019.....	96
¿Son los mismos desafíos para ambas jurisdicciones?.....	96
Algunas ideas para seguir pensando.....	100
BIBLIOGRAFÍA.....	106

INDICE DE FIGURAS Y CUADROS

Figura 1: Esquema comparativo de las prácticas de calificación	40
Figura 2: Matriz de Validez de Messick.....	47
Figura 3: De la edad media a los colegios jesuitas.....	54
Figura 4: Condiciones históricas de desarrollo de las prácticas de evaluación.....	55
Figura 5: Matriz progresiva de validez.....	64
Figura 6. Ejemplos de las categorías de la matriz de validez y sus comentarios.....	68
Cuadro 1 Escenario 1.....	71
Cuadro 2 Escenario 2.....	73
Cuadro 3 Escenario Cuadro 3.....	75
Cuadro 4 Escenario 4.....	77
Cuadro 5 Escenario 5.....	79
Cuadro 6 Escenario 6.....	81
Cuadro 7 Escenario 7.....	83
Cuadro 8 Escenario 8.....	84
Cuadro 9 Escenario 9	87
Cuadro 10: Fundamentos de validez comparativo por asignaturas.....	88
Cuadro 11: Fundamentos de validez total.....	89
Cuadro 12: Respuestas y antigüedad docente.....	91
Cuadro 13: Fundamentos de validez por jurisdicción.	99

INTRODUCCIÓN

Hoy está instalada la cuestión de la revisión de la evaluación de los aprendizajes desde diversas corrientes y con distintas motivaciones. Tanto desde aquellas perspectivas que buscan dar cuenta del proceso individual que recorre el alumno en la apropiación del conocimiento (nivel micro), como las miradas que ponen el foco en la *accountability* para medir resultados que permitan diseñar políticas nacionales y globales para la mejora de la educación (nivel macro); todos ponen a la evaluación y sus prácticas como eje central de los procesos de innovación y mejora de la educación (Shepard, 2006).

Los movimientos de reformas educativas, en muchos países ya son vistos como parte de la rutina o ciclos de cambios que se suceden. Aunque respondan a distinto tenor, orientación política y criterio pedagógico, sin embargo, todos se han orientado al currículum (así las políticas del “*Common Core*” en Estados Unidos u otras revisiones generales de la escuela secundaria como en Francia) y, muy raramente, han abordado las prácticas de evaluación y calificación.

La evaluación en el aula constituye uno de los pilares de la cultura escolar dominante. Y, sin embargo, en general la ignoramos. No suele ser objeto de investigación y tiene un lugar absolutamente secundario en la formación de los docentes. Es un elefante invisible. (Ravela, Picaroni y Loureiro, 2017: 31)

En nuestro país, las dos leyes más importantes de las últimas décadas a nivel educativo (Ley Federal de Educación en 1993 y Ley Nacional de Educación en 2006) han procurado ampliar el derecho a la educación con la extensión de la obligatoriedad y lograr una oferta acorde que permita hacer efectivo ese derecho para todos los niños, niñas y adolescentes. En la escuela media, donde vamos a concentrar nuestro análisis, ha habido diversas reformas curriculares que han buscado, de diverso modo, dar cuenta de la inadecuación de los saberes que ofrecía una enseñanza tradicional y- salvo algunos casos como la escuela técnica- totalmente alejada del mundo del trabajo y de profundizar algunas competencias claves que se necesitan en los estudios superiores.

Las normativas que han llevado adelante la aplicación de las leyes referidas poco y nada han abordado las cuestiones más profundas de la gramática escolar de la enseñanza secundaria. Entre ellas, por ejemplo, la asignación de horas cátedra como esquema laboral o las prácticas de evaluación y calificación. Parece haber allí un foco muy resistente al cambio y que, a pesar de intentar algunas modificaciones, terminan todas subsumidas en lo que perciben los docentes y la comunidad educativa en general (padres y alumnos) acerca de la importancia de las notas, qué valor tienen y qué relación tienen con el efectivo desempeño académico del estudiante.

Dado el proceso de implementación de la Ley Nacional de Educación, que pone como objetivo prioritario la inclusión con calidad, se hace necesaria una nueva escuela secundaria que efectivamente garantice a todos el ingreso, permanencia, movilidad y egreso (CFE 93/09, art. 4). Lo que los documentos y normativas más recientes suelen llamar “trayectorias escolares continuas y completas” (CFE 93/09, art. 3).

Dentro de dicho proceso de resignificación de un nivel educativo que no fue obligatorio sino optativo y selectivo (recién se extiende la obligatoriedad con la ley Federal de Educación en 1993), es fundamental revisar las prácticas de evaluación, acreditación y promoción de los alumnos.

El “Marco para la Escuela Secundaria Federal 2030” pone foco en los procesos que se activaron con las leyes mencionadas y, a su vez, en los límites de sus logros:

Los datos objetivos mencionados – sólo 5 de cada 10 jóvenes terminan la secundaria y los datos de Aprender 2016- son los que revelan que las políticas educativas implementadas hasta el momento no han sido suficientemente potentes para reconvertir el funcionamiento originario de la escuela secundaria, de una escuela tradicionalmente selectiva a una escuela inclusiva, abierta y de calidad. (Marco para la Implementación de la Escuela Secundaria Federal 2030, 2017: 5)

El Documento “Orientaciones para la organización pedagógica e institucional de la educación secundaria obligatoria”, aprobado por la mencionada resolución 93/09 ya proponía hace 10 años una seria revisión de las prácticas de evaluación y calificación:

“Los procedimientos vigentes recurren a una escala numérica para valorar la complejidad de los procesos de aprendizaje, que resulta insuficiente para dar cuenta de los mismos (...). Resulta inadecuado simplificar una decisión de vital importancia para la trayectoria escolar de los jóvenes a una cuestión de operaciones con cantidades que representan momentos particulares y no el desarrollo completo de un proceso. La calificación que el estudiante obtiene en cada período o año escolar no debe limitarse a una operación cuantitativa como es el promedio aritmético” (CFE 93/09, art. 138).

Estas resoluciones procuraban orientar la mejora de las condiciones de acreditación y promoción, pero no fue hasta el año 2018 con la implementación de “Secundaria del Futuro 2030” en las primeras escuelas (llamadas pioneras o promotoras) que comienza a ponerse en marcha en el Ciclo Básico del nivel secundario una profunda transformación que abarca los aspectos organizativos de la escuela media como el nombramiento de profesores por cargo y, nuestro tema de interés, la revisión de los Regímenes Académicos.

En la formación básica del docente, en la comunidad académica, en los cursos de “capacitación” de la más diversa índole que se brindan por vía del sistema tanto de gestión estatal como privado ha habido una creciente preocupación por la evaluación, como se

evidencia en la producción bibliográfica en nuestro medio y en otros países. Lo que se llama en la jerga de habla inglesa *classroom assessment* (para diferenciarla de otras instancias y situaciones de evaluación que utiliza el sistema a nivel nacional o internacional) ha sido abordado procurando incorporar otra mirada sobre el error, una mayor conexión entre el estilo de enseñanza y la evaluación, una evaluación que pueda brindar mayor información al alumno de cómo mejorar su aprendizaje (evaluación formativa) y distintas modalidades que podrían englobarse en lo que se ha dado en llamar evaluación formativa.

Así lo vemos en las investigaciones de Anijovich (2010,2011), Camilloni (1998, 2004, 2010), Cappelletti (2010), Litwin (1998), – entre otros- y en los materiales que producen éstas y otras autoras más orientadas al docente y la aplicación de ese marco teórico a las prácticas de evaluación en el aula.

Sin embargo, es muy escasa la investigación sobre calificación y casi inexistente la capacitación, materiales o instancias de reflexión orientada a los que aplican la normativa sobre calificación: los docentes. Pareciera que la percepción generalizada refleja una visión de la calificación como algo absolutamente normado, carente de toda posibilidad de gestión del docente, el departamento o el colegio. Puede que el director, la familia de un estudiante o incluso el mismo docente se pregunten cómo evalúa, qué dispositivo ofrece e incluso “cómo corrige” un profesor. Pero pensar cómo se utiliza el sistema de calificación, si éste cumple realmente con los objetivos educativos de una nota o si ésta refleja o no el aprendizaje logrado por un estudiante, parecen ser cuestionamientos ajenos a la mayoría de los actores del sistema en el debate académico de nuestro país. Lo que no quiere decir que no se converse en las salas de profesores, no genere conflictos en la cotidianidad de la escuela o, incluso, sea uno de los temas más recurrentes cuando la conversación informal en otros ámbitos se involucra en temas educativos en nuestro país y, aparentemente, en todo aquel que transita o ha transitado las aulas del nivel secundario en los más diversos países del mundo.

No extraña, pues, lo que comentan Guskey y Brookhart en su reciente revisión sobre las investigaciones en calificación y sus prácticas: “Basta sacar el tema de cómo califican los docentes si quieres comenzar una discusión sobre algún tema educativo” (Guskey y Brookhart, 2019:1).

Por otra parte, “la forma en que los docentes elaboran las notas dista mucho de ser evidente para todos los alumnos, y aún para sus padres. (...) funcionan como índices que anuncian que todo va bien, que la carrera académica sigue su curso, que el alumno trabaja con normalidad; o que hay que empezar a preocuparse, que está presente la amenaza de fracaso escolar” (Perrenoud, P., 2001: 144).

Aunque se implementen otras estrategias de evaluación bien fundadas en el proceso -las rúbricas, por ejemplo, que van ganando más y más terreno en el aula y ahora de implementación obligatoria en Secundaria Federal 2030-, al momento de calificar mucho de todo lo bueno y completo hecho en clave de evaluación formativa termina estrellándose en la construcción de una única nota trimestral que es fruto de un promedio matemático (media) que surge de notas de producciones, tests, pruebas, exposiciones o incluso alguna nota de concepto con una misma o diferente ponderación para la conformación de la calificación trimestral. Así pues, las calificaciones “tienen un efecto tan poderoso que suelen operar de manera muy efectiva, en reversa, como determinantes de todas las prácticas pedagógicas” (Camilloni, 1998: 133).

Guskey y Brookhart (2019) recuerdan que

...la calificación es uno de los debates más calientes de los temas de educación y las prácticas de calificación están ampliamente basadas en la tradición, el instinto o la historia o filosofía personal. Para ser efectivas, las políticas de calificación y las prácticas deben basarse en evidencia confiable de investigación. (Guskey y Brookhart, 2019: 3)

Estos mismos autores expresan su preocupación porque, en el concierto de una necesidad de revisar las prácticas de calificación, -en particular impulsados por la implementación del *Common Core* en Estados Unidos y la *Standards Based Assessment* – muchos docentes no acceden a la investigación, sino que se informan y capacitan a través de blogs, libros de difusión, páginas de Facebook y hasta tutoriales en Youtube. “Algunos de esos autores de libros sobre calificación no citan una sola investigación en calificación y muchos bloggers sobre notas hacen lo mismo (...)” (Guskey y Brookhart, 2019: 2), lo que los motivó al equipo que había realizado la revisión de los 100 años de investigación en calificación a producir una versión ampliada de la revisión para que pueda llegar a los docentes y directivos implicados en las iniciativas de reforma de la normativa de calificación.

Las notas son parte de la vida cotidiana de un estudiante, ya desde los últimos años de la educación primaria, pero especialmente en el secundario. Va a ser el inmediato motivo de consulta de parte de cualquier persona que le pregunte cómo le está yendo en la escuela. ¿Está claro para qué sirven las notas? ¿Tienen para todos los docentes la misma función y significado? ¿Qué importancia tienen en la “gramática” de la escuela media? ¿De dónde salen las notas? ¿El alumno “se saca” notas o el profesor “se las pone”? Así como se evidencia poca formación docente en evaluación, ¿se prepara el futuro profesor para calificar? ¿Cómo?

Es llamativo, a riesgo de ser reiterativo, que una práctica tan central en la escuela sea tan poco abordada y que – incluso en los cursos de grado o posgrado en evaluación– no se la considere.

Transparentar los criterios de validez subyacentes a las decisiones que toman los profesores a la hora de realizar el cierre del trimestre, se convierte en un conocimiento valioso para la organización de los nuevos regímenes académicos que se están revisando e implementando en el marco de las resoluciones del CFE ya citadas, y desde 2018 el Marco General y los Escenarios Posibles de adecuación del Régimen Académico que el Ministerio de Educación de la Nación ha propuesto. Así mismo permitirá diseñar políticas y normativas que faciliten al docente establecer más claramente los logros equivalentes de cada estudiante en un diseño curricular que está volcando su orientación hacia las competencias para el mundo globalizado del siglo XXI.

Así, pues, esta investigación procurará acercarse a las prácticas de calificación de los docentes de la escuela secundaria, tomando una muestra de profesores de Lengua, Matemáticas y Biología que se desempeñan en las jurisdicciones de Ciudad de Buenos Aires y Provincia de Buenos Aires, aclarando que no pertenecen a las llamadas escuelas promotoras que están implementando el nuevo régimen académico para el Ciclo Básico del Nivel Secundario. Creemos que, muy especialmente en estos años, serán las calificaciones un tema central y los regímenes académicos un instrumento fundamental para abordar los objetivos aún no alcanzados de la Ley de Educación Nacional.

Llamativamente, pareciera que Guskey y Brookhart hablan de nuestro país cuando afirman:

La calificación es típicamente el último elemento que se encara en los esfuerzos de reforma educativa. Muchas iniciativas de reforma comienzan con - la articulación

de los objetivos del diseño curricular o los estándares -esto es, qué deben aprender y ser capaces de hacer los estudiantes como resultado de sus experiencias en la escuela-. Luego, consideramos qué evidencia refleja mejor ese aprendizaje y comenzamos a analizar cuestiones asociadas con la evaluación de los estudiantes. Después de esto, nos focalizamos en la calidad de la enseñanza y cómo ayudar mejor a los alumnos a alcanzar los objetivos o estándares establecidos. Finalmente, giramos nuestra mirada hacia las calificaciones -la cuestión de comunicar efectivamente los resultados de evaluación y otras medidas de aprendizaje a los estudiantes, los padres y las familias-. Nos hacemos cargo de la calificación al final y siempre con cierto rechazo porque cambiar las políticas y prácticas de calificación significa desafiar una de las tradiciones educativas de más larga data. (Guskey y Brookhart, 2019: 2).

Espero sepan disculpar la cita extensa, pero difícilmente pudiera encontrar palabras más claras y, especialmente, autorizadas para dejar clara la oportunidad y pertinencia de este primer abordaje descriptivo de las calificaciones en la escuela secundaria.



Universidad de
San Andrés

CAPÍTULO 1

LA INVESTIGACIÓN EN EVALUACIÓN Y CALIFICACIÓN AYER Y HOY

¿Qué aportaron más de 100 años de investigación?

En el medio anglófono hay abundante literatura respecto de las prácticas de calificación, tal como surge de la revisión que coordinara en 2016, Susan Brookhart para la convocatoria que hizo AERA (American Educational Research Association), en ocasión de su centenario, a publicar síntesis de los temas más significativos para los educadores en *Review of Educational Research*. Con otros de los más prestigiosos especialistas en calificación: Thomas Guskey, Alex Bowers y James Mc. Millan -entre otros- repasó, analizó y clasificó lo más importante en 100 años de investigación en calificación (Brookhart, Guskey, Bowers, McMillan, Smith, J., Smith, L., Stevens y Welsh, 2016)

Todos los estudios sobre las calificaciones o notas, aunque se hayan acercado al tema desde distintos puntos de vista: experiencias de corrección múltiple de exámenes, confiabilidad, validez, percepción de los docentes de sus propias prácticas, representaciones de los estudiantes acerca de las notas, su sentido y función, etc. siempre buscan contestar a la misma pregunta que, por lo tanto, se presenta siempre compleja y contextual y, como otros temas de las ciencias sociales y de la educación elusivo por su carácter dinámico. “La pregunta central que subyace en todos estos estudios es: “¿Qué significan las notas?”. Es, esencialmente una pregunta por la validez (Kane, 2006; Messick, 1989)”. (Brookhart *et al*, 2016: 804).

La centralidad de las notas en la vida de los estudiantes es un tema crucial que preocupó desde su misma aparición, sin embargo, al momento de encarar cambios y transformaciones de los sistemas educativos -y así lo hemos visto desde la década del 90 en Argentina- suele ser uno de los últimos temas sobre los que se comienzan a plantear e implementar estrategias de innovación y mejora. “Nos hacemos cargo de las calificaciones a último momento y siempre con cierto rechazo porque, cambiar las políticas y prácticas de calificación, significa desafiar unas de las tradiciones de más larga data en la educación” (Guskey y Brookhart, 2019: 2).

1. Los primeros estudios

El sistema educativo se encontró, - casi al mismo tiempo en que se comenzaba a

masificar y necesitó una forma rápida y eficiente de indicar el nivel de aprendizaje de un estudiante, con un problema concreto: “hay una gran variación en las notas que asignan los docentes a sus estudiantes” (Brookhart *et al*, 2016: 806).

Basada en la teoría de la curva normal. Edgeworth -en 1888- aplicó la teoría de los errores realizando experiencias de corrección múltiple: se entrega el mismo examen a distintos correctores para comparar qué calificación le otorga cada docente. Describió tres tipos de fuentes de error -que bien podrían ser los mismos para hoy, más de 130 años después-: el azar, cuestiones ligadas al examen en general como severidad, laxitud y velocidad y el hecho de puntuar las preguntas del examen como representativas del dominio del tema (Brookhart *et al*, 2016). “La idea de que múltiples factores llevan a notas que no son confiables fue un enorme paso adelante. Les dio a los educadores algunas pistas acerca de qué podían hacer frente al problema” (Brookhart y Guskey, 2019: 17-18). Desafíos a la confiabilidad, que hoy se siguen enfrentando y se busca mejorar. Por ello, el primer grupo que abordó esta revisión general fueron los estudios sobre confiabilidad que investigaban la corrección de producciones individuales, generalmente exámenes, no calificaciones que combinan otras notas como en el caso de las trimestrales que se publican en los boletines. Algunos estudios sólo constataron sin más el hecho de que realmente existía este fenómeno de enorme variabilidad en la forma de puntuar y corregir, otros procuraron indagar en los factores que contribuían a ello: la escala de calificación fue objeto de revisión y de propuestas alternativas que llevaron, por ejemplo, a que lentamente se fueran imponiendo las notas con letras en todos los Estados Unidos para la década del 40.

Fueron muy difundidos los estudios realizados por Daniel Starch y Richard Elliot (1912, 1913a, 1913b; Starch, 1913, 1915) que encontraron como fuente principal de las variaciones a los errores de los docentes y al uso de diferentes criterios. Llevaron adelante una muy prolífica serie de investigaciones que implicaban la corrección de pruebas de inglés, matemáticas e historia. Hacia 1915 propusieron distintas estrategias que ayudaran con la confiabilidad de las notas. Algunas parecen muy valiosas aún hoy: estudiar cómo califican y discutirlo en reunión de departamento, acordar en su seno cómo puntuar determinados errores comunes y -una idea que aún tiene fuerte impacto en la educación secundaria- que las notas habrían de estar distribuidas según la curva normal u otra distribución estandarizada. Había encontrado rangos de diferencia de entre 40 y 50 puntos sobre 100 en las materias estudiadas (Cureton, 1971). Más tarde, presentaron un pequeño

estudio que mostró (1915) cómo ajustar las notas a 5 categorías basadas en la distribución normal, lo que reducía su variabilidad de 4.2 puntos a 2.8 (Brookhart y Guskey, 2019).

Unos años más tarde, ya a fines de la década del 20, todo un movimiento en Francia influenciado por la psicología y la psicometría va a dar lugar a una nueva disciplina que estudia los exámenes: la docimología. En la década del 30, el estudio experimental de calificación de exámenes de Baccalauréat que llevaron adelante Laugier y Weinberg (1930) concluye, por ejemplo, que los métodos de evaluación utilizados corrientemente son totalmente insuficientes ya que las notas asignadas a un mismo trabajo podrían representar, según los correctores hasta un rango de 13 puntos sobre una escala de 20 (Cardinet, 1989; Martin, 2002). También la confiabilidad comenzaba a preocupar seriamente a la comunidad francesa dada la importancia que tenían los exámenes de selección profesional y el tradicional -y aún vigente- Baccalauréat como examen final del nivel secundario y requisito obligatorio para el ingreso a la universidad.

2. La investigación en calificación: temas y enfoques de ayer y hoy.

En el relevamiento que tomamos como referencia (Brookhart *et al*, 2016) se destacan otros temas, como la relación entre las notas y otros resultados educativos en 3 áreas: relación entre notas y tests de inteligencia estandarizados; las calificaciones como medidas multidimensionales de conocimiento académico, compromiso y perseverancia y, la relación entre las notas y otros indicadores de la educación como los que miden el abandono o la graduación del nivel secundario. “Estos estudios ofrecen evidencia de validez predictiva bajo el supuesto de que las notas miden el éxito escolar” (Brookhart *et al*, 2016: 821). Si el estudiante tuvo tales o cuales notas en primer año, por ejemplo, qué posibilidades tiene de abandonar el colegio y/o de graduarse en tiempo y forma.

En la indagación por la correlación entre las notas y los tests de inteligencia se han encontrado algunas moderadas, pero marcadamente consistentes en los estudios. Bowers (2011) identificó un 25 % de la variación en las notas atribuible a rasgos que miden los tests, en tanto que Swineford señaló en un estudio sobre las calificaciones en la escuela intermedia y secundaria que “Los datos muestran claramente que las notas asignadas por los docentes en esta escuela son medidas confiables de algo, pero aparentemente no hay acuerdo acerca de que habría de ser ese algo” (1947: 517).

Entre las investigaciones sobre la multidimensionalidad de las notas, se halló -por

ejemplo, en los trabajos de Kelly (2008) y Bowers (2009 y 2011)- que el compromiso es parte importante de la composición de las notas (Brookhart *et al*, 2016).

Un área que se desarrolla cada vez más en la última década es la de los estudios que se dedican a la evaluación basada en estándares (SBG *standards based assessment*), enfoque que enfatiza la comunicación del progreso del estudiante en relación con distintos niveles de aprendizaje que describen el desempeño utilizando categorías ordinales e implican informar separadamente aquellos aspectos que tienen que ver con los hábitos de trabajo, la conducta y la responsabilidad (Brookhart, 2011; O'Connor, 2009; Guskey, 2011). Sin embargo, aunque hay mucha literatura de divulgación, capacitación y recomendaciones, aún son muy pocos los estudios empíricos que investiguen las prácticas concretas y la validez de la evaluación basada en estándares (Welsh, 2019). Dado que es una herramienta de calificación que utiliza en gran medida rúbricas con un enfoque criterial y formativo, las investigaciones que aborden SBG podrían compararse en el futuro con aquellas que se lleven adelante a partir de las experiencias de las escuelas promotoras del programa “Escuela Secundaria Federal 2030” que -desde 2018- se viene poniendo en marcha en distintas jurisdicciones de la Argentina.

La validez de las calificaciones en la percepción del docente

En el marco de los objetivos de esta investigación, nos interesa recuperar la reseña de trabajos que realizó Brookhart (1994) y que su revisión más reciente (Brookhart *et al*, 2016) ubica bajo el grupo de estudios dedicados a la percepción de los docentes acerca de las calificaciones y las prácticas de calificación. En este grupo de 19 trabajos empíricos acerca de las opiniones y creencias que tienen los profesores acerca de sus propias prácticas de calificación aparecen 5 cuestiones principales:

1. El elemento más determinante en las calificaciones que utilizan los docentes como medida de logros de aprendizaje son los tests (o pruebas).
2. Los profesores consideran que es importante ser justos a la hora de calificar.
3. Utilizan para ello muchas fuentes de información.
4. Incorporan el esfuerzo en las notas.
5. Les preocupa comunicar claramente a los estudiantes qué evaluarán y cómo serán calificados los estudiantes (Brookhart *et al*, 2016: 825)

“La conclusión es que las notas típicamente representan una mezcla de múltiples factores que valoran los docentes. Reconocen el rol importante del esfuerzo en la motivación y los logros (Aronson, 2008; Cizek *et al.*, 1995; Cross & Frary, 1999; Duncan & Noonan, 2007; Guskey, 2002,

2009a; Imperial, 2011; S. Kelly, 2008; Liu, 2008b; McMillan, 2001; McMillan et al., 2002; McMillan & Lawson, 2001; McMillan & Nash, 2000; Randall & Engelhard, 2009, 2010; Russell & Austin, 2010; Sun & Cheng, 2013; Svennberg et al., 2014; Troug & Friedman, 1996; Yesbeck, 2011). Diferencian facilitadores académicos (McMillan, 2001:25) como esfuerzo, habilidad, mejora, hábitos de trabajo, atención, participación, a los que incorporan como relevantes para las notas. (Brookhart et al, 2016: 834).

Destacamos algunas consideraciones que los propios autores consideran relevantes:

Yesbeck (2011) estudia los factores académicos y no-académicos que se presentan en las prácticas de calificación. Señala que, a lo largo de la investigación, uno de los temas recurrentes es la falta de desarrollo profesional para los docentes en medición -cómo crear pruebas que midan lo que pretenden medir (temas de validez) y cómo, consiguientemente, calificar esas mediciones. Va a concluir que los docentes mezclan variables en el fenómeno que describimos en el marco teórico como *hodgepodge grading* (Brookhart, 1991 y 1993).

Caroline Hummel Wiley (2011) se propone estudiar los perfiles de las prácticas de calificación de los profesores. No sólo para determinar qué tipo de factores consideran, sino también, examinar la congruencia entre las diversas fuentes de evidencia para calificar y sus respectivas creencias. Lo novedoso es que este estudio indaga cómo esas prácticas impactan en alumnos concretos. En la misma línea que Yesbeck (2011) y Brookhart (1993) subraya la mezcla de factores, pero, en su caso llama a las prácticas de calificación un “proceso confuso -fuzzy- que incluye factores no académicos en las notas” (2011: 14). Imperial (2011), que implementó una encuesta de 63 ítems cuidadosamente desarrollada, encontró que aproximadamente el 50 % de los profesores indicaba que las notas representan solamente rendimiento académico, en tanto que un 78 % incluía la tarea, 57 % el esfuerzo y 70 % la participación.

Matthew Bailey (2012), que procuró establecer una correlación entre las percepciones que tienen los docentes de secundaria de sus prácticas de calificación y la percepción del docente respecto de la motivación de los estudiantes, comenta que los resultados le indicaron que las docentes mujeres enfatizaban más la conducta del estudiante que los docentes varones y que los profesores de ciencias le dan menos importancia al esfuerzo que los docentes varones de ciencias sociales (McMillan, 2019: 90).

Los trabajos de McMillan (1998, 2001) y sus propias revisiones de investigaciones (2001, 2013 y 2019) describen también estos fenómenos incorporando la metáfora de

facilitadores académicos *-academic enablers-* para describir todas esas conductas y actitudes que, si bien no son resultados de producciones específicas de los estudiantes, sin embargo, son incluidos en la nota por los docentes porque precisamente son conductas que consideran fundamentales para poder obtener mejores logros académicos. En este sentido McMillan termina abogando por cierta práctica de calificación idiosincrática e individualizada con rasgos de *hodgepodge*, como ya lo habían planteado Cross y Frairy (1996) al comprobar que esa forma de calificar era auspiciada tanto por los estudiantes como por los docentes, a pesar de la recomendación de los especialistas de poner las notas sólo desde las producciones académicas.

Sintetiza McMillan, con bastante claridad, al finalizar la extensa revisión que publica este año 2019:

En lugar de denigrar la calificación “hodgepodge”, sería más útil comprenderla mejor y ayudar a los profesores a utilizar tanto factores académicos como no-académicos de modo de que sean bien pensados, transparentes y justos para aumentar la motivación y el aprendizaje de los estudiantes (McMillan, 2019: 107)

Llama la atención la vigencia del interés por desentrañar los fundamentos de las decisiones que toman los profesores al calificar, en la encuesta Teacher’s Grading Practices Survey llevada adelante por Thomas Guskey y Laura Link (2018) en 5 distritos escolares del sudeste de los Estados Unidos. Allí identificaron diferencias entre los docentes del nivel primario y del nivel secundario en el énfasis otorgan al esfuerzo como parte de la calificación

Si bien las referiremos frecuentemente en nuestro capítulo metodológico ya que utilizamos el mismo test de escenarios, ha resultado sumamente interesante el abordaje propuesto por Susan Brookhart (1993) de docentes en formación de posgrado para investigar sus prácticas de calificación, especialmente el significado y valor que le asignan las notas. A un dispositivo diseñado originalmente por Manke y Loyd en 1990, la investigadora incorporó una pregunta por los fundamentos que sustentan las decisiones de calificación que toman los profesores, de acuerdo con la situación que les planteó cada escenario del test. Analizó dichos comentarios que explicitan los criterios desde la matriz de validez progresiva de Messick (1989b) y encontró que el docente está más preocupado por las consecuencias del uso que se les dará a las notas que comunica a su alumno que en una adecuada interpretación de lo que las calificaciones significan en sí:

Los docentes no siguen las prácticas de calificación recomendadas en parte porque hay un conflicto

entre las prácticas recomendadas, que se concentran en la interpretación y el sentido de las notas y sus preocupaciones acerca del uso de las calificaciones (Brookhart, 1993: 139).

Más recientemente, Youyi Sun y Liying Cheng (2014), también interesadas por indagar el significado y valor que les asignan a las calificaciones, han replicado esta investigación con docentes de inglés en China utilizando algunos de los escenarios que aplicó Brookhart (1993). Identificaron -entre las preocupaciones de los profesores- la apelación a lo que es justo y beneficioso para los estudiantes. Por otra parte, destacan el carácter contextual de esta práctica al destacar que “los docentes hacen estos juicios basados en los valores que sostienen acerca de las calificaciones, que están a su vez asociadas con los múltiples roles que cumplen en el contexto sociocultural” (Sun y Cheng, 2014: 340).

3. Otros idiomas ¿las mismas preocupaciones?

En idioma francés, tal como reseña Lucie Mottier López (2010) la investigación tiene como referentes a Perrenoud (Suiza) De Ketele (Bélgica) y Merle (Francia). El movimiento “contra la constante macabra” de André Antibi (2011) se está instalando en Francia que vive un contexto de profunda revisión de la escuela media. Gana fuerza y promueve todo un espacio de indagación específicamente orientado a las notas y su utilización. De entre estos autores uno de los aspectos que más aborda Perrenoud (1989) es la construcción del éxito y el fracaso escolar a partir de las prácticas de evaluación y calificación que los profesores llevan adelante.

Es Pierre Merle (2007, 2014a, 2014b) quien más se ha volcado a la descripción del proceso de fabricación, negociación y construcción de las calificaciones en un análisis sociológico que daríamos en llamar intersubjetivo. Llama a las prácticas de calificación negociación o arreglos *-arrangements-*. No obstante, también nos permitimos comparar una de sus metáforas conceptuales (Camilloni, 2014). El uso que hace -con Perrenoud (1984)- de la metáfora del *bricolage* para caracterizar las prácticas de calificación que analiza exhaustivamente como fabricación nos remite a las utilizadas para caracterizar estas peculiares prácticas de calificación que los docentes de habla inglesa llaman *hodgepodge* (Brookhart, 1991 y 1993; Yezbeck 2011), *kitchen sink* (Cizek, Fitzgerald y Rachor, 1995-1996) o *fuzzy process* (Hummel Wiley, 2011).

Si nos centramos en las prácticas de evaluación de los aprendizajes que los docentes llevan adelante en el aula, se encuentra abundante producción académica en la Argentina

y América Latina. En Argentina muchos autores han reflexionado sobre la evaluación, sus características, sus problemas, la introducción de nuevas prácticas e instrumentos de evaluación alternativa e incluso su revisión crítica en los distintos niveles de enseñanza (Anijovich, 2010 y 2017; Camilloni, 1998; González, 2011; Bonvecchio de Aruani, 2011; Celman, 1998; Elola, 2010; Litwin, 2009; Palou de Maté, 1998). Otros han relevado las representaciones de alumnos, profesores en formación y docentes acerca de la evaluación y la calificación (Del Corral, 2007; Morejón, 2004; Pruzzo de di Pego, 1999).

Pero, si bien hay abundante investigación acerca de la evaluación en el aula, “la cantidad de trabajos empíricos existentes decrece considerablemente a la hora de describir las prácticas pedagógicas de evaluación en escuelas concretas” (Macha, 2006: 2), como así también es muy escasa la indagación acerca de las prácticas de calificación propiamente dichas en Argentina, y, en el mejor de los casos aparecen junto a otros trabajos sobre evaluación.

Nos parece importante señalar - de entre sus numerosos y profundos trabajos sobre evaluación- los aportes de Camilloni a la reflexión sobre los sistemas de calificación y regímenes de promoción (1998) de los que afirma -en un medio que no publica ni reflexiona sobre las prácticas de calificación- que “no acompañan las transformaciones de las prácticas de la enseñanza (...) obligando a los docentes a una permanente adaptación a los moldes pedagógicos que los procesos de renovación o reforma pretenden reemplazar” (1998: 133). Sus aportes teóricos al concepto de validez (1998 y 2010) de la evaluación y la calificación brindan al debate académico una fundamentación de la que generalmente carece la bibliografía que suele estar más orientada a las recomendaciones prácticas que necesita o requiere el profesor para el diseño de instrumentos de evaluación. Desde la didáctica, la reflexión que propone Alicia Camilloni sobre las metáforas conceptuales en la construcción del discurso pedagógico (2014), en tanto “orientan la comprensión y la construcción de significado, sustentan la construcción de conocimiento e influyen significativamente sobre ésta” (2014: 20) nos ha permitido detenernos en nuestro estudio sobre el uso que han realizado de dichas metáforas distintos investigadores en su tarea de describir y comprender mejor las prácticas de calificación y su producto que es la nota. Hemos reseñado dichas expresiones en este estado del arte y las iremos retomando en distintos momentos de la investigación.

Otra autora argentina, Susana Celman (1998), también ha utilizado una metáfora

conceptual muy sugerente al describir a la nota como un “signo opaco”. Al requerirle mayor información a la autora, ya que no desarrolla el concepto en el ensayo referido, nos comunicó por mail que no ha escrito más al respecto. No obstante, comenta en ese correo personal, que “alude a su dificultad de comunicar significados porque sólo indica un lugar en una escala. La nota no abre la posibilidad de comprender ni las razones de su atribución ni los procesos que determinaron su adjudicación. Al no comunicar el proceso, no permite visualizar relaciones significativas, ni trabajar sobre el error. Parecen primar en ella las funciones de jerarquización y selección” (Celman, 2015) más que las de brindar información relevante sobre su aprendizaje al estudiante, al docente, a sus padres y demás interesados.

Recientemente Ravela, Piccaroni y Loureiro (2017) desde Uruguay y México han publicado un material que, si bien está más orientado a los docentes, incorpora muchas investigaciones realizadas por sus autores e integra la calificación en el proceso de enseñanza, aprendizaje y evaluación proponiendo una reformulación formativa de la evaluación sumativa. Dedicar un capítulo específico a las prácticas de calificación en el que realizan una profunda mirada comparativa de reglamentos de evaluación y calificación de varios países de América Latina y señalan algunos aspectos clave de la calificación en la práctica docente: la falta de significado sustantivo de las notas en términos de aprendizaje, su matematización y el uso del promedio como juicio de valor de la persona del estudiante. Recogiendo en distintas investigaciones comentarios de los docentes sobre sus propias prácticas identifican los mecanismos de su construcción centrados en la inclusión de muchas calificaciones, la calificación del esfuerzo y la concepción de la nota como pago al alumno. “Los alumnos toman en serio sólo que se califica” (Ravela *et al*, 2017: 232) recogen los autores de una entrevista a una docente. Hacen una propuesta de trabajo para los docentes con el notable mérito de incorporar mucho material de investigación empírica sin caer en la receta didáctica de aplicación inmediata sino promoviendo una reflexión sobre las prácticas que parte de otros colegas cuyas palabras han recogido estos investigadores.

Llama la atención que también el material coordinado por Brookhart y Guskey (2019) haya surgido, según su propio testimonio, del pedido de la ASCD (Association for Supervision and Curriculum Development) de difundir entre los docentes la revisión de las investigaciones sobre calificación que los mismos autores habían publicado en la *Review of Educational Research* en 2016, más orientada hacia los investigadores y

académicos. A pesar de que Susan Brookhart, Thomas Guskey, Robert Marzano y Pedro Ravela están muy vinculados a sesiones presenciales y virtuales de capacitación del más variado tipo, y realizan numerosas publicaciones con orientaciones prácticas para el docente, parece que hay una preocupación por conectar cada vez más y mejor la investigación con el aula y el aula con la investigación. Es una invitación -el investigar las propias prácticas y compartirlas con sus colegas - que también necesitamos promover entre los profesores para comprender más y realizar mejor nuestra tarea de enseñar, evaluar y calificar.



Universidad de
San Andrés

CAPÍTULO 2

LA EVALUACIÓN Y LA CALIFICACIÓN EN EL PROCESO DE ENSEÑANZA Y APRENDIZAJE

Para presentar el enfoque teórico desde el que abordamos nuestro objeto de investigación -las notas que asignan los docentes a los alumnos en la clase y, particularmente, sus prácticas de calificación en ocasión del cierre del período trimestral- se hace necesario ubicarlo en el contexto teórico-conceptual en que se define y utiliza: la evaluación en el marco del proceso de enseñanza y aprendizaje de una clase y su posterior comunicación (reporting) a los padres y al sistema en informes y boletines.

Para ello presentaremos nuestra concepción de evaluación educativa, una somera clasificación que hace particular énfasis en la distinción y complementariedad entre evaluación sumativa y formativa, para, finalmente, abordar las cuestiones centrales que hacen a las notas, sus prácticas de elaboración por parte de los profesores y las cuestiones, problemáticas y desafíos que hoy las atraviesan.

1. La evaluación: una práctica social

“Evaluar es establecer un juicio acerca de una cosa o persona. El evaluador no es un mero analizador de datos sino alguien que juzga, que toma una decisión” (Gvirtz, S. – Palamidessi, M; 2008: 241). Michael Scriven (2013) aclara que es un proceso cognitivo y, aunque el contenido de las afirmaciones evaluativas sea extremadamente dependiente del contexto, no significa que su significado no sea claro, cuestión de gustos idiosincráticos o esencialmente subjetivos. Es, pues, un juicio de valor, la apreciación acerca de la calidad de un objeto que puede ser una obra de arte, o un instrumento de la vida cotidiana o el trabajo realizado y visibilizado en un producto material. Siempre evaluar es elucidar el valor de algo realizado por alguien. Por eso la evaluación es vivida como un cierto riesgo o amenaza cuando está vinculada a algo que nosotros hemos realizado. Así pues, en el Diccionario de la Real Academia Española, se define el verbo “evaluar” como “señalar el valor de algo”, “estimar, apreciar, calcular”. No se trata, entonces, solamente de un saber específico del ámbito educativo, “evaluar es una actividad esencial y natural para el ser humano” (Ravela, Picaroni y Loureiro, 2017: 31). Lo hacemos todo el tiempo, en forma explícita e implícita, teniendo plena consciencia de ello o de modo inadvertido. Cada vez que tomamos una decisión a nivel individual o social realizamos algún tipo de evaluación previa basada en el análisis de la información

que hemos recogido, como lo señalan Ravela *et al.* (2017). Por ello, el ya referido Michael Scriven, la denomina una transdisciplina que aporta herramientas a las restantes (2013: 36).

Evaluación educativa

La tercera y última acepción de la RAE está orientada hacia este aspecto que ha atravesado la sociedad desde hace tres siglos y, al pasar todos por la escuela, ha llegado casi a ser significado excluyente de evaluar: “Estimar los conocimientos, aptitudes y rendimiento de los alumnos”.

Podemos definir la evaluación educativa como el proceso de recolectar e interpretar evidencia del progreso del estudiante que permita realizar un juicio fundado de lo que el estudiante o el grupo conoce en relación a los objetivos de aprendizaje (National Research Council, 2001 en Moss, 2013). Podemos ver, pues, a la evaluación como construcción de conocimiento, un proceso de aprendizaje (Sverdlick, 2012), un proceso de “diálogo, comprensión y mejora” dirá Santos Guerra (1995: 34) ya indicando en dicha afirmación que el conocimiento que brinde la evaluación no se agota en el proceso, sino que encuentra su finalidad en la retroalimentación que provee.

Por otra parte, la evaluación en educación sigue siempre una misma lógica asentada sobre bases comunes (Camilloni, 2004; Ravela *et al.*, 2017), a saber:

1. Un objetivo: queremos decidir si estamos logrando lo que nos proponemos en nuestras clases o cómo ayudar a los estudiantes a avanzar o si un estudiante concreto está o no en condiciones de aprobar el curso;
2. Necesitamos establecer alguna afirmación valorativa sobre la situación del grupo en general o de estudiantes específicos en relación con los aprendizajes que queremos promover: para definir o reorientar cursos de acción posibles en la propuesta didáctica o bien, en otros casos para establecer las calificaciones que constituyen un juicio de valor sobre su desempeño;
3. Tomamos en consideración ciertas dimensiones o criterios relevantes y damos importancia relativa a cada una de ellas.
4. Recogemos alguna información o evidencia empírica sobre lo que cada estudiante puede hacer en relación con algún atributo, un aspecto, un rasgo, un conocimiento; es decir, aquellos aprendizajes que hemos establecido como relevantes.

5. A partir del análisis de esa información recogida realizamos un “trabajo de coligación bastante complejo a través del cual establecemos una conclusión evaluativa” (Ravela *et al*, 2017: 36) o juicio de valor (Camilloni, 2004: 6). En algunos casos será una nota (evaluación sumativa o certificativa) y en otros cómo continuar el proceso educativo y ayudar a los estudiantes a avanzar en el aprendizaje (evaluación formativa).

Estas bases van a estar orientadas por la finalidad que tenga el proceso evaluativo emprendido, de aquí que “es importante recordar en forma constante que la evaluación no tiene un fin en sí misma” (Anijovich, 2011: 13). Es dicho objetivo el que va a determinar cómo lleva adelante el docente este proceso de evaluación, no los instrumentos, dispositivos o métodos de recolección de la información (Moss, 2013).

El tipo de evaluación y sus fines van a orientar la elección de un tipo de información y un instrumento utilizado para recogerla, la forma de medición que realicemos comparativamente bien en relación a una escala o a una producción de otras personas (por ejemplo del grupo clase), o nos llevará a presentar un determinado modelo de logro y el respectivo juicio de valor que construimos a modo interpretativo de los datos obtenidos codificado en los signos que cada sistema educativo provea a tal fin.

2. Tipos de evaluación

Evaluaciones externas estandarizadas

Si pensamos en el objeto de la evaluación educativa, podemos hablar de evaluación de los desempeños de estudiantes, de los docentes, de los directivos, de la institución escolar, del sistema. Cuando nos ubicamos en diferentes escalas y consideramos al estudiante como parte de un colectivo pequeño que llamamos curso o clase, hablamos de toda una disciplina que denominamos evaluación en el aula (classroom assessment) y cuando se evalúa el desempeño de estudiantes a nivel distrital, jurisdiccional, nacional e internacional se plantea -con muchos años de historia y evolución- toda otra disciplina de evaluaciones externas estandarizadas a gran escala (large scale assessment / external testing). En algunos sistemas educativos esto casi ha convivido desde el comienzo de su constitución como tales. Así en Estados Unidos desde los albores del siglo XX se considera que han operado dos sistemas de evaluación en forma simultánea (Brookhart, 2013a) : el juicio del docente expresado en la calificación de evaluaciones sumativas informadas en los boletines (de gran impacto individual en el estudiante) y, por otro lado,

las pruebas externas para investigación, evaluación de programas e instituciones (de gran impacto para las escuelas) que fueron aumentando y se siguen complejizando progresivamente al involucrarse otros actores como el estado federal y las pruebas internacionales (PISA y TIMSS principalmente).

En América Latina la irrupción de las evaluaciones externas es mucho más reciente. Los organismos de evaluación nacionales y jurisdiccionales han surgido contemporáneamente a la adhesión de los países -con sus altos y bajos- a las pruebas internacionales como PISA. Axel Rivas señala que:

“El auge de las evaluaciones de la calidad educativa comenzó en los años noventa pero explotó en los dos mil. Fue parte de un gran proceso de reingeniería de los sistemas educativos. Las políticas de descentralización y autonomía escolar llevaban en su vientre a las políticas de evaluación (Weiler, 1996). Fue el inicio de la era educativa de “devolución de responsabilidades”: el Estado central les decía a los niveles intermedios de gobierno y a las escuelas que se alejaba de la gestión para evaluar los resultados. Si la descentralización fue la gran fuerza de cambio en los noventa, la evaluación tuvo impacto real en los dos mil” (Rivas, A., 2015: 91).

Desde ya que el impacto de las evaluaciones externas centralizadas fue muy variado en las distintas experiencias de los países de la región, siendo Argentina y Uruguay precisamente casos que el autor llama de “baja exposición pública y escaso impacto en la vida de escuelas y docentes” (Rivas, A., 2015: 94). Richard Stiggins (1993) habla de dos disciplinas de evaluación educativa claramente diferentes, que incluso lleva a replantear la enorme cantidad de recursos destinados a las evaluaciones estandarizadas a gran escala frente a la evaluación cotidiana en las clases:

“Estos datos del sistema parecen detenerse en la puerta del aula. Los docentes no los utilizan. No los encuentran útiles. Creo que esto se debe a que estos sistemas fracasan en reconocer las diferencias entre la evaluación y la toma de decisiones del docente en la realidad del aula y la evaluación que hacen los políticos y administradores del sistema para las decisiones que enfrentan (Stiggins, R., 1993: 4).

Podemos decir entonces, a partir del testimonio de Stiggins (1993) que, cuando comenzaban las pruebas y organismos nacionales de evaluación en América Latina, ya se objetaba su interacción con la realidad concreta de los estudiantes en las aulas por parte de los más importantes especialistas en países que llevaban casi un siglo de “práctica simultánea”.

La evaluación en el aula

Como señalábamos antes, la evaluación no tiene un fin en sí misma. Por lo tanto, el tipo de evaluación va a estar determinada por el uso que pretendamos hacer y la decisión que tengamos que tomar a partir de la información que nos brindan las evidencias

que ha arrojado.

Podemos tomar la somera y clásica clasificación de evaluación que nos propone Merle (2014): sumativa, certificativa, formativa y diagnóstica. Sus finalidades se configuran en- sus características, prácticas e instrumentos para recoger la información. Las calificaciones -objeto específico de esta investigación- son propias de las evaluaciones sumativas y certificativas, que parecen tener mucho más en cuenta las autoridades educativas por la necesidad de acreditar saberes para la promoción y la emisión de títulos (función certificativa). Incluso hoy, constata Merle en la misma conferencia, muchos docentes desconocen la evaluación formativa y todo el marco teórico que la sustenta. Puede que lleguen a incorporar algunos de sus instrumentos o modalidades, pero sin ver la distinción -y según algunos la complementariedad- entre los dos grandes campos de la evaluación en el aula: la evaluación sumativa y la evaluación formativa. Es la misma tensión que refiere Perrenoud (2008) cuando se pregunta si la evaluación está al servicio de la clasificación o del aprendizaje.

Uno de los aspectos que parece ser una convicción común a todos los investigadores y especialistas en evaluación es la necesidad de una adecuada integración de la evaluación, la calificación y la información (reporting) al proceso de enseñanza y aprendizaje (Camilloni, 1998). Es parte integrante y fundamental no sólo como instancia que da información al estudiante y al profesor para retroalimentar el proceso que cada uno de los actores viene desarrollando (evaluación formativa) o al finalizar un determinado período o curso para brindar una certificación determinada y/o brindar a posibles instancias de estudios superiores un cierto reporte de cuál ha sido el desempeño de ese estudiante (evaluación sumativa y certificativa). Además de brindar información, una evaluación plenamente integrada a la enseñanza y al aprendizaje produce conocimiento y forma en lo mismo que está haciendo: la habilidad de evaluar, ponderar, discriminar, valorar, criticar, opinar, razonar, fundamentar, decidir, elegir. “Esta actitud evaluadora, que se aprende, es parte del proceso educativo que, como tal, es continuamente formativo” (Álvarez Méndez, 1996 en Celman, 1998:37).

Evaluación formativa

La evaluación formativa, término propuesto en 1967 por Michael Scriven, procura, además de indicar al alumno sus aciertos y errores, brindarle retroalimentación vinculada

a criterios claros de desempeño y estrategias de mejoramiento (en Shepard, 2006).

Paul Black y Dylan Wiliam (2009), dos de los especialistas que más han sistematizado los fundamentos teóricos de la evaluación formativa la definen como “un proceso en el que se recaba información con el fin de revisar y modificar la enseñanza y el aprendizaje en función de las necesidades de los alumnos y las expectativas de logro para alcanzar. La información que se busca tiene relación con las representaciones mentales del alumno y con las estrategias que utiliza para lograr un resultado” (en Anijovich, 2010: 10). Realiza una plena articulación de la evaluación con el aprendizaje y la enseñanza ya que apunta a ofrecer al alumno una devolución (feedback) que le permita identificar dónde está en su proceso de aprendizaje, a dónde espera la propuesta pedagógica que llegue y cómo puede recorrer el camino que aún le queda para poder llegar a las metas esperadas (Stiggins, Arter, Chappuis, J y Chappuis, S., 2004)

De las múltiples definiciones que recoge Camilloni (2004) del propio Scriven (1967), Airasian (1971) y Perrenoud (1999) entre otros, propone dos características que parecen comunes:

- La contemporaneidad con los procesos de enseñanza y de aprendizaje.
- Intención de que la información permita mejorar los procesos evaluados.

Respecto de este punto Perrenoud (2008) va a focalizar en el aporte que hace la evaluación formativa a la formación de los procesos de autorregulación de los aprendizajes por parte del alumno y otros, como Stiggins (2002), van a ir perfilando más claramente la función principal de la evaluación formativa al llamarla, precisamente, evaluación para el aprendizaje (assessment for learning) en tanto que la evaluación sumativa sería la evaluación del aprendizaje (assessment of learning) en tanto cierra el proceso de enseñanza y aprendizaje al valorar el producto.

Susan Brookhart la define por contraste:

“La evaluación formativa implica reunir y comunicar información para su uso en el desarrollo de conocimiento y habilidades, y la evaluación sumativa implica información reunida e informada para su uso en juzgar los resultados de ese desarrollo. Como se dice: “Cuando el cocinero prueba la sopa, eso es evaluación formativa. Cuando el cliente prueba la sopa, eso es evaluación sumativa” (2009: 45).

Por otra parte, algunos especialistas siguen viendo una “falta de alineamiento en las definiciones de evaluación formativa y sumativa y sus relaciones mutuas, así como una brecha teórica que deriva en una divergencia entre teoría y práctica” según refieren Aitken

y McKean (2016:26).

La evaluación formativa tiene un valor en sí misma, más allá de su necesaria articulación con la sumativa y, en general, con todo el proceso de enseñanza y aprendizaje. Las estrategias e instrumentos que pone en juego (entre ellos las rúbricas) activan el desarrollo de capacidades de revisión, reflexión sobre las propias producciones y qué recursos ha utilizado el estudiante para ese proceso de aprendizaje.

Evaluación sumativa

La evaluación sumativa apunta a brindar un registro de los resultados globales del proceso de aprendizaje de determinados conocimientos y habilidades de un estudiante en un área específica y en un momento determinado (Harlen, 2004; Brookhart, 2009).

Shepard advierte que “la evaluación sumativa y la calificación constituyen una seria amenaza para los objetivos de aprendizaje declarados en la evaluación formativa” (2006: 30). Más allá del específico diseño de instrumentos o propuestas, la evaluación se convierte en formativa o sumativa de acuerdo al uso que se le da. Si el estudiante puede utilizar la información que se le devuelve para regular su aprendizaje, es claramente formativa. Si cierra ya dicho proceso porque terminó el año escolar o la asignatura o el itinerario pedagógico se abre a temas muy diferentes, la evaluación será sumativa: sintetiza en un código de calificación que procura informar acerca de resultados logrados a los distintos actores: autoridades del sistema, directivos, profesores, padres y estudiantes.

El cierre del trimestre y del año son las instancias de evaluación sumativa en las que se condensan y evidencian todas estas concepciones de evaluación que subyacen a las decisiones que toman los docentes. Perrenoud también señala el límite y disociación que se produce entre evaluación y calificación (o evaluación formativa y sumativa), cuando “la evaluación formativa se choca con la calificación y la evaluación tradicional que llamamos normativa. Aunque los compromisos tradicionales de evaluación van influyendo menos, la evaluación formativa no exime a los profesores de poner notas o de redactar informes cuya función no es ayudar a los alumnos a aprender sino de informar a los padres o a la administración escolar” (1989:6).

Si bien las notas (números, letras u otros símbolos utilizados para informar la

calidad de desempeño) son consideradas generalmente evaluación sumativa, el proceso de calificar, especialmente cuando se califican producciones, “porta una gran cantidad de información que los estudiantes pueden utilizar para mejorar” (Brookhart 2009: 45), es decir, con una finalidad formativa.

En el prólogo del manual editado por James McMillan en SAGE y dedicado a la investigación de la evaluación en clase, Shepard (2013) plantea qué condiciones deben darse para una adecuada articulación entre evaluación formativa y sumativa que oriente la agenda de investigación y experimentación:

“La evaluación formativa y la evaluación sumativa son compatibles si se enfocan en las mismas ricas, desafiantes y auténticas metas de aprendizaje y si el feedback que se realiza en medio de la enseñanza conduce a una comprensión profunda y un desempeño mejorado en las tareas finales sumativas” (Shepard, L, 2013: xxi).

Pone por aviso a investigadores y reformadores que quieran desarrollar prácticas de calificación más acordes con la investigación en aprendizaje y motivación, que deben tener en cuenta los siguientes obstáculos:

1. Están trabajando contra creencias fuertemente arraigadas entre docentes y estudiantes, y “cuanto mayores sean los estudiantes, más explícitas habrán de ser las negociaciones para volver a focalizar el esfuerzo en el aprendizaje” (Shepard, L, 2013: xxi).
2. Los docentes están hoy en día fuertemente presionados a mantener informados a los padres con sistemas informáticos en tiempo real. Esto implica, con frecuencia, que la nota ya se coloque, aunque el aprendizaje no haya finalizado.
3. En un ambiente de evaluaciones externas de variado tipo, frecuentemente se distorsiona el sentido del feedback, poniendo el énfasis en los dos o tres puntos que el estudiante tiene que mejorar para aprobar en lugar de qué aspectos fundamentales del tema necesita aún comprender mejor. (Shepard, L, 2013).

A pesar de esta extendida reflexión y profundización acerca de la evaluación formativa e incluso del uso de la calificación y la evaluación sumativa como insumo formativo para la mejora del aprendizaje y de las estrategias de enseñanza, sin embargo sigue utilizándose en el lenguaje cotidiano evaluación como sinónimo de calificación.

Dice López Pastor que:

“... lo verdaderamente grave no es lo generalizada que está dicha confusión terminológica, sino que las prácticas evaluativas (formativas, educativas) han desaparecido bajo la preeminencia y dominio de las prácticas calificativas (...) al final lo que cuenta es la evaluación sumativa y acreditadora (...)

la nota” (López Pastor, 2005: 2).

En nuestro país, recientemente afirmaba rotundamente Ingrid Sverdlick que, desde una lógica administrativa que parece ser prioritaria en el sistema “la evaluación queda subsumida en la calificación” (2012: 150). Por ello se hace fundamental analizar las prácticas de calificación, ya que son parte definitoria de toda evaluación y, en una mirada integral, de todo aprendizaje. Si no se incorporan nuevas miradas que nos permitan revisar la forma en que el profesor pone notas, difícilmente puedan consolidarse estrategias de evaluación formativa que ayuden a un mejor aprendizaje. Veamos de qué se trata esta práctica en el marco de la teoría de la evaluación que hemos delineado.

3. Las prácticas de calificación¹

Antes de explicitar nuestro posicionamiento respecto del concepto de nota que articula nuestra tarea, compartimos algunas expresiones que, más que definir la nota, nos arrojan sobre la problemática y paradojas que la atraviesan tanto en lo que es como en las formas de su elaboración y producción que reflejan sus prácticas.

Paul Dressel expresaba lo siguiente:

“Una calificación puede ser considerada como un reporte inadecuado de un juicio impreciso, realizado por un juez sesgado y variable, acerca del grado en que un estudiante ha logrado un nivel indefinido de dominio, de una proporción desconocida de un material indefinido”. (Dressel, P., 1957: 6)

Más allá de su fina ironía, plantea los mismos problemas que hoy enfrenta la nota y las prácticas que conducen a su elaboración, aunque varios de esos aspectos se estén abordando más o menos sistemáticamente tanto desde la academia como desde las políticas educativas:

- Reporte inadecuado (el producto: la nota). La función principal de las notas que es informar es confusa, por tanto, se manifiesta como inadecuada.
- Juicio impreciso (la acción interpretativa del docente): el juicio de valor fruto de la interpretación de la evidencia que el docente ha hecho es vago.

¹ A lo largo de la presente investigación utilizaremos como sinónimos intercambiables los vocablos “nota” y “calificación”, sea que nos refiramos a la asignación de una nota a una determinada producción (por ejemplo, una prueba) o estemos refiriéndonos a la informada en un boletín de calificaciones como “nota trimestral”.

- Grado de logro del estudiante de un nivel indefinido (qué ha evaluado el docente): parece mostrar una absoluta falta de transparencia y claridad en el nivel esperado ni tampoco qué logró o no logró respecto de dicho nivel (proporción).
- Material indefinido: parece remitir a que se desconocen los contenidos o habilidades de los que debía dar cuenta y/o los objetivos que se pretendían. Especialmente en las notas trimestrales que sintetizan información muy diversa.

De allí que surjan varias metáforas descriptivas de las prácticas de calificación que dan cuenta de que la función de informar se ve seriamente afectada por su forma de producción:

- “*hodgepodge grading*” -revoltijo, mezclanza-, que acuña Brookhart (1991) y luego es ampliamente retomada por otros especialistas (Cross, 1993; Yezbeck, 2011; Sun y Cheng, 2014; Guskey y Link, 2018).
- “*kitchen sink*” -fregadero, pileta de la cocina- (Cizek, Fitzgerald y Rachor, 1995-1996; Bowers, 2011).
- “*fuzzy*” - confuso, vago, borroso- (Hummel-Wiley, 2011).
- “signo opaco” (Celman, S., 1998).
- “*grading conflation*” -amalgama de notas- (Guskey y Link, 2018: 2).

Muchos investigadores y docentes dan cuenta de la incomodidad que genera evaluar y, en particular, calificar: “No nos gusta evaluar y no sabemos hacerlo del todo bien, pero calificamos o asignamos puntajes la mayor parte del tiempo” (Ravela *et al*, 2017: 27). “En un mundo perfecto no habría notas, al menos tal y como las conocemos hoy” (Brookhart, 2009: 2). Como lo indicaba Shepard (2006) respecto de la evaluación sumativa, “la calificación “mata” la evaluación formativa” (Ravela *et al*, 2017: 210), porque se transforma en el objetivo del estudiante que adopta una actitud pasiva respecto del aprendizaje. Si las notas son el gran objetivo detrás del que van los estudiantes, el aprendizaje pasa a un segundo plano. No bien recibe la nota, el aprendizaje se detiene, “puede no gustarnos, pero la investigación muestra que esto es una característica relativamente estable de cómo trabaja la mente humana” (William, 2011: 123). La nota cierra, clausura, finaliza un proceso de aprendizaje más allá de cómo haya sido. Si el docente considera que es importante que el estudiante vuelva a revisar su producción para mejorarla, corregir algunos errores y seguir adelante con el proceso, se debe evitar asignarle -al menos por ahora- una calificación.

Por eso, es fundamental que la finalidad de la nota sea clara y transparente ya que tiene una enorme carga de determinación sobre todo el proceso de enseñanza y aprendizaje del que forma parte indisoluble -como lo destacamos previamente- la evaluación. En definitiva, pese a los más fuertes detractores de la nota que

Algunos autores tienden a definir a la nota como “la posición en un continuum de calidad, dominio, intensidad o valor. Este continuum puede ser expresado de forma numérica, con letras o utilizando descriptores verbales” (Anderson, 2018: 3). En la evaluación tradicional -según Merle (2007)- tiende a asociarse a la calificación como la ubicación del estudiante en una escala, para lo que, simplemente, se aplica un baremo. Otros parecen reforzar, en sus definiciones la codificación que refleja un determinado nivel de aprendizaje: “una nota es esencialmente una representación simbólica del nivel de logro alcanzado por un alumno” (Sadler, 2009: 1).

Con una mirada más completa, Brookhart y Guskey, en la más reciente revisión de investigaciones sobre calificación, las definen como

“los símbolos asignados a producciones individuales de los estudiantes o medidas combinadas del desempeño de los estudiantes, creadas para los boletines y otros documentos sumativos. Pueden ser letras, números, figuras u otro conjunto de descriptores para designar distintos niveles de desempeño” (2019:1).

Estamos hablando, entonces, de las notas de una prueba, un trabajo práctico, una exposición oral o cualquier otro desempeño del estudiante y, otra práctica específica: la nota que se coloca en un boletín al fin del trimestre, fruto de algún tipo de operación con las notas de esas producciones y que arroja la calificación trimestral como resultado. Esencialmente “las notas no son nada más ni nada menos que etiquetas pegadas a diferentes niveles de desempeño del estudiante (...)” (Guskey & Brookhart, 2019: 217).

Las calificaciones responden a la pregunta que siempre hacen los estudiantes: “¿Cómo me fue?” o -aún más orientado al mero reporte numérico- “¿Qué me saqué?”. Cabría preguntarnos si a dicha expresión subyacen supuestos de azar o lotería, metáfora que utiliza Bruno Suchaut para referirse a la “inexactitud y arbitrariedad” de las calificaciones (2008: 1). Como si las notas fueran parte de un sino fatal fruto de alguna mágica intervención, de algún modo reflejada también en la expresión de algunos docentes: “te quedó” tal nota o “el promedio te dio...” o -incluso- “el sistema te cerró así...” como si la responsabilidad de la calificación estuviera en otro lugar y no tuviese

nada que ver con las decisiones, las calificaciones parciales y las operaciones con ellas que ha realizado el profesor.

Stiggins, Friesbie y Griswold (1989) presentan la agenda de investigación en calificación, lo que nos ayuda a identificar cuáles son los ejes principales de análisis que necesitamos abordar para una aproximación al tema. Consideran que es vital estudiar los procedimientos específicos que siguen los docentes cuando califican al fin de un período trimestral, cómo deciden qué procedimientos van a utilizar y -finalmente y en forma indisoluble- qué calidad tendrán las notas que surgen como fruto de esos procedimientos. Es vital no sólo identificar qué es o qué significa una determinada nota para un docente (Anderson, 2018; Brookhart, 1993; Sun y Cheng, 2014), sino que es vital analizar por qué un profesor elige un determinado modo de elaborar la calificación trimestral e informarla utilizando determinado código de entre aquellos disponibles en la normativa que lo encuadre en la jurisdicción educativa en la que se desempeñe (Guskey y Link, 2018: 13).

Anderson (2018) señala que las críticas al modo en que se asignan las notas han arreciado en los últimos años. En distintos países se reiteran estos cuestionamientos pese a las diferencias en la normativa, escala e implicancias para la trayectoria del alumno (Dilon, 2018; Moinet, 2018; Sverdllick, 2012; López Pastor, 2004). Sin embargo, se trata de un fenómeno paradójico. Pese a todas las críticas y las falencias que han puesto en evidencia las investigaciones desde hace más de 100 años (ver capítulo 1 sobre el estado del arte de la investigación en el área), las calificaciones siguen allí en el sistema casi de modo inalterado. Su ausencia en la formación inicial de los docentes que reflejan los investigadores (Brookhart, 1993; O'Connor, 2010; Merle, 2014) a la par que cierta "ilusión de transparencia" (Bourdieu, 2002) que naturaliza la omnipresencialidad de las notas y su lugar central en la vida de los estudiantes y las escuelas, parecen reflejar un auténtico obstáculo epistemológico (Bachelard, 1971). Algunos podrían pensar que se trata de un dispositivo nodal de la gramática escolar (Tyack y Cuban, 1995) que, de transformarse sustancialmente o directamente desaparecer -los dos grandes grupos de recomendaciones que proponen los expertos- alterarían los propios -cimientos de la escuela y el sistema educativo. Un gran impulsor de la evaluación formativa nos advierte reiteradamente que calificar más no es calificar mejor (William, D., 2011). Sin embargo, reconoce que: "...no nos podemos deshacer de las notas porque todo el sistema Americano de admisión al College depende de ellas" (William, D., 2011:123). La práctica de admisión a las universidades en nuestro país casi no tiene en cuenta las calificaciones de

la escuela media -sólo en algunas universidades privadas les da facilidades para una beca o ayuda económica ó evitar la realización de cursillos breves de ingreso obligatorios para los estudiantes con nota inferior a un determinado promedio-. Por lo tanto, ese aspecto que señalan los autores norteamericanos como barrera insalvable para la desaparición de las notas, no estaría operando en nuestro sistema educativo. Schneider y Hutt señalan otro foco de resistencia al desplazamiento de la nota como insumo central del sistema: “parece que hay una tensión inevitable en la escuela moderna entre aquello que promueve el aprendizaje y lo que permite que funcione un sistema masivo” (2014: 217).

Reflejan esta universalidad de la experiencia de la inadecuación y límite de las notas como indicadores de aprendizaje a la vez que su persistencia, las reflexiones que comparte Weiss, desde Suiza cuando señala que la nota numérica se ha impuesto y es muy difícil de remover porque es cómoda, fácil de comprender, es considerada justa, precisa e indiscutible (1996). Todos ven en ella – dice- una “precisión matemática bienvenida en una escuela tan conflictiva y compleja” (Weiss, 1996: 1), permite tomar decisiones administrativas, estimular el trabajo del alumno y manejar (o dominar) la gestión de la clase.

Cuando se estudia la historia de las calificaciones, como lo hacen estos autores se descubre que las calificaciones fueron masificándose y tomando mayor protagonismo a medida que creció exponencialmente el sistema educativo, especialmente el nivel secundario y superior.

El ámbito académico plantea fuertes objeciones a la nota, tanto en el medio anglófono con Alfie Kohn ya desde los 90 (1994;1999;2010), como en los sectores académicos de habla francesa con André Antibi (2012) y su “movimiento contra la constante macabra”, Perrenoud (2008) o Maulini (2004). Sus críticas y eventuales propuestas van desde la prudente transformación de sus prácticas e interpretación hasta una eventual abolición de las calificaciones, al menos como las conocemos hasta hoy.

Vemos también cierta paulatina y aún discreta irrupción en el debate público con las charlas TED de Mark Barnes, de la estudiante Sofía Camussi de Argentina, o el surgimiento de páginas específicas para educadores como panote.org que dan cuenta de un tema que aún sigue generando preguntas y problemas no resueltos.

O’Connor, en un enfoque pragmático parte del hecho de que las mediciones que más utilizan los docentes y las escuelas son las calificaciones, entonces, “las mediciones

que necesitan las escuelas realmente son notas bien hechas” (2009:1). Con ese mismo tipo de realismo Brookhart va a apuntar siempre, siendo plenamente consciente de las dificultades en la validez (1991), la distorsión en la motivación (2003) y otras consecuencias negativas del sistema, a mejorar la formación en evaluación y calificación (1993; 2009; 2017) porque las demandas actuales de escolarización están lejos de permitirnos la utopía de prescindir de las calificaciones, “por ello es importante aprender a calificar bien” (2009: 4).

Digamos que las notas están ahí y parecen ser consideradas parte indisoluble de la escuela secundaria, por lo tanto más allá de su posible remoción o cambio esencial en la escuela del futuro se hace necesario revisar su producción, su interpretación y el uso que le damos a la información dentro y fuera de la escuela.

Las notas: su finalidad

Es una pregunta central que va a estar íntimamente imbricada en la forma en la que dicha nota se asigna, elabora, construye, fabrica o negocia. Claro, estos verbos que podrían dar cuenta de las prácticas de calificación ya dicen algo diferente sobre su significado al describir la acción que “da a luz” una nota con una palabra diferente. Asimismo, el posicionamiento teórico pedagógico que tomemos respecto de la nota va a estar marcado por su finalidad particular dentro del amplio marco de la evaluación en general y de la evaluación sumativa, en particular.

¿Para qué están las notas? Para comunicar (reporting) a distintos actores del sistema: el mismo estudiante y sus padres (podríamos hablar de un primer círculo de actores, junto al docente que califica) (Brookhart,2009) luego otros actores o interesados (stakeholders) que forman un segundo círculo: los responsables del sistema educativo a nivel municipal-distrital, provincial y federal -que tienen preocupaciones políticas respecto de la rendición de cuentas de la calidad del sistema (accountability) y finalmente lo que daríamos en llamar un tercer círculo de interesados que necesitan eventualmente información del desempeño de los estudiantes para su inserción, luego de su egreso de la escuela secundaria, al ámbito laboral o a estudios superiores (Withney, Culligan y Brooksher, 2006; Schneider y Hutt, 2014).

En la historia del surgimiento y evolución de las calificaciones, esta finalidad de la demanda externa a la escuela va a ir presionando hacia una cada vez mayor uniformidad, en demerito de una adecuada precisión en la información que brinda sobre el desempeño del estudiante. Como lo abordaremos luego, la pretensión de fortalecer la validez predictiva en las calificaciones suele afectar seriamente la validez de construcción, erosionando el significado de las notas. Se necesita no solamente que las calificaciones “tengan un significado similar entre las distintas clases sino, idealmente, entre las distintas escuelas” (Schneider *et al*, 2014:209).

Allí tenemos un primer desafío. ¿Son el único sistema para comunicar a padres y alumnos acerca del logro de sus objetivos de aprendizaje? (Brookhart, 2009). Sin duda no, y es cada vez más frecuente complementar dicha síntesis de la calificación con feedback que pueda incorporar el estudiante para la mejora de su proceso de aprendizaje (en ese sentido dándole una orientación formativa a esa nota) o bien distintos descriptores que procuren enriquecer y, me atrevería decir, desentrañar algo del misterio que está detrás de ese número más allá de su supuesta construcción matemática como promedio de las distintas instancias o producciones calificadas durante el período. Sin embargo, son la única información oficial y, por lo tanto, la que tiene impacto en las trayectorias de cada alumno. Aprueba ahora o sigue ajustando su aprendizaje y siendo examinado (diciembre y/o febrero con distintos formatos) o bien es reprobado debiendo, eventualmente, repetir el curso o año. Estas son las características comunes de la normativa de calificación vigente en la escuela media en Argentina.

Hasta allí podríamos hablar de una finalidad primaria que luego condensa en otras que diversifican su alcance por las consecuencias que implican en las decisiones que toman los otros interesados “fuera del aula” ya mencionados.

Sin embargo, a veces solapadamente y en otras ocasiones favorecido explícita o implícitamente por los regímenes de evaluación, acreditación y promoción, otras finalidades se entremezclan.

Guskey (2011) señala que fuertes tradiciones compartidas por docentes y directivos, padres y alumnos obstaculizan cualquier reforma del sistema de calificación y - aún cuando algunas de ellas estén desalentadas por los académicos o incluso sean contrarias a las normativas vigentes- operan fuertemente en las prácticas de calificación. Desde estas perspectivas con fuerte arraigo, las notas:

- Deben proveer las bases para diferenciar a los estudiantes. Esta es la función de selección y clasificación, que Perrenoud (2009) denomina construcción del éxito y del fracaso escolar. “Los estudiantes que muestran un talento superior reciben notas altas, en tanto que aquellos que despliegan un talento inferior reciben notas más bajas” (Guskey, 2011: 17). Esto lleva a maximizar las diferencias para lograr la mayor variabilidad posible, optando, por ejemplo, por una escala de 100 grados de calidad y/o construyendo un dispositivo que sea insensible a la enseñanza, entre otros efectos que produce establecer esta finalidad para las calificaciones.
- La distribución de las notas en un grupo-clase deben tomar la forma de una curva normal en campana, dado que los puntajes en un test de inteligencia -relacionada indudablemente con el aprendizaje- tienden a formar dicho dibujo. Este supuesto ya fue claramente desestimado por la investigación (Bloom, 1976) porque dicha correlación precisamente cambia o debería cambiar su “forma natural”, por las intervenciones del docente en la enseñanza. Sin embargo, aún hoy, muchos docentes corrigen las pruebas de sus estudiantes en una primera etapa sin asignarles una nota y luego los califican comenzando por las mejores producciones a las que les asignan la mayor nota en la escala, distribuyendo el resto en forma proporcional, con un criterio que se aproxima al que Meyer (1908) diseñó para la Universidad de Missouri: 3 % excelentes, 3 % desaprobados, un 50 % en el nivel promedio , un 22 % con calificaciones por sobre el promedio (superior) y otro 22 % apenas por debajo del promedio (inferior). (Cureton, 1971) Otras distribuciones se ensayaron y/o funcionan como discursos legitimados en la sala de profesores: “si un docente tiene más del 10 o 15 % de alumnos desaprobados, el que falla es el profesor”.
- Deben asignarse en referencia a la ubicación del estudiante respecto de su grupo de clase. El efecto es el mismo que en el supuesto anterior pero aquí ya no está fundado en una supuesta distribución científica de las capacidades en la naturaleza sino que lisa y llanamente explicita la necesidad de elaborar rankings de alumnos. Es una creencia de larga data

-desde las escuelas jesuitas del siglo XVI que realizan compuestas individuales y grupales organizando las clases según el modelo de las legiones romanas (Maulini, O, 1996)- sostenida en el aparente valor de la competencia entre alumnos como un aliciente al aprendizaje (Anderson, 2018). Daña las relaciones en la escuela (Krumboltz & Yeh, 1996), dificulta el aprendizaje grupal y la cooperación entre los alumnos. Además “se transforma en una valoración de la persona que inmoviliza (entre aquellos a los que no les va bien)”. (Ravela *et al*, 2017: 266) y a los alumnos que suelen tener mejores calificaciones los desalienta a tomar cursos de mayor dificultad o emprender proyectos que implican un mayor desafío cognitivo (Brookhart, 2009:4).

Estos 3 presupuestos que hemos explicado se ubican en lo que se da en llamar el enfoque normativo de las calificaciones (Ravela *et al*, 2017). ¿Qué aporta la información de este tipo de prácticas de calificación? Poco y nada, como lo ratifica Guskey: “...las notas que se basan en la posición del estudiante entre sus compañeros de clase no nos dicen nada acerca de la calidad de lo que han aprendido” (2011: 18).

- Las notas bajas motivan a los estudiantes a esforzarse más. Actuarían, según este criterio extendido, como motivación extrínseca. En realidad, las bajas notas desalientan a los alumnos llevándolos incluso a abandonar el esfuerzo y la dedicación -poca o mucha que hayan puesto de su parte hasta ese momento- a esa asignatura.
- Los estudiantes son calificados con una sola nota en cada asignatura. Calificación en el boletín que pretende ser una síntesis de informaciones generalmente diferentes pero que deben quedar subsumidas en un único número. Suelen combinar pruebas de control, ensayos, informes, proyectos y evaluaciones de fin de unidad con cumplimiento de las tareas, puntualidad, participación, esfuerzo, etc. Esto lleva a uno de los fenómenos que han dado en llamar Brookhart (1991;1993;2009) y otros “hodgepodge grading” (revoltijo, mezclanza).

Dado que la operación de síntesis generalmente está establecida como un promedio, algunos procuran jerarquizar estos distintos componentes estableciendo una ponderación diferente a cada una de los aspectos,

producciones y dimensiones del desempeño del alumno de los que la nota única trimestral daría cuenta.

La expresión de este juicio global en una única nota termina siendo una valoración de la persona del estudiante más que una información relativa a las habilidades o saberes que hubiere adquirido (Ravela *et al.*, 2017).

Esta información que brindan -sobre todo la nota trimestral- va a ser utilizada por la escuela (los docentes, los directivos y/u otros responsables del sistema) para certificar una determinada asignatura al fin de su cursado, la aprobación de un determinado año y la respectiva promoción al siguiente o, al final de un ciclo del sistema, la obtención de un título determinado. En algunos países (Estados Unidos, Canadá, Francia, entre otros) esta información además puede determinar la ubicación del alumno en una clase avanzada o de recuperación, su postulación a una beca y/o la posibilidad de ingresar a una determinada institución de estudios superiores, en tanto el promedio general (GPA en Estados Unidos) puede ser uno de los factores que le permitan ser admitido o no en la universidad de su preferencia (Withney *et al.*, 2006).

También, sigue Susan Brookhart en su manual de calificación (2009), entre los propósitos secundarios de la calificación está el proveer a los docentes con información para (re) diseñar sus planificaciones, a los directivos y actores políticos del sistema (inspectores, secretarios, ministros) para evaluar distintos proyectos o diseños curriculares. ¿Por qué objetivos o finalidades secundarias? No porque sean de una menor importancia sino porque para las decisiones que tienen que tomar esos actores la información que brindan las notas es insuficiente. Ya sea para evaluar una determinada estrategia didáctica o enfoque de enseñanza nuevo o la reorganización de objetivos en la secuencia de aprendizaje anual o de ciclo.

En nuestro país no existe un examen final “posterior” que legitime o clasifique a los egresados del nivel secundario para estudios superiores o el ámbito laboral. Por lo tanto, podríamos decir que las notas y el título con su certificado analítico son las únicas y últimas informaciones que brinda la educación secundaria en nuestro medio. Sin embargo, también esos actores del sistema: las universidades, los departamentos de recursos humanos que reclutan trabajadores -con variada frecuencia y eventualmente escaso impacto- suelen recurrir al promedio general del nivel o incluso - aunque raramente- analizan el certificado analítico que evidencia en su valor simbólico y legal la

necesidad de registrar las calificaciones que ha obtenido en cada asignatura en cada año del nivel secundario. Desde fines de los 90 y hasta, aproximadamente 2010, en el nivel secundario de la Ciudad Autónoma de Buenos Aires se modificaron los documentos de certificación. Se eliminaba el certificado analítico que indicaba la nota final de cada asignatura en cada uno de los años del nivel medio y sólo se colocaba el promedio final general de todo el nivel. No faltaban estudiantes que solicitaban una copia certificada del libro matriz con sus calificaciones materia por materia, bien por interés propio y/o porque se lo solicitaban del ámbito laboral o académico para ingreso y/o el acceso a algún programa desarrollo profesional, intercambio interuniversitario beca.

¿Cómo se hacen o fabrican o producen? Las prácticas de calificación

Además de definir qué son las notas, es importante identificar el proceso que las produce. Incluso, en inglés, la palabra “grade” puede remitir tanto al sustantivo (la nota, la calificación) como al verbo (to grade: calificar). En tanto que las prácticas de calificación (grading) refieren el proceso de resumir el logro de un estudiante en un símbolo o número (Brookhart, 2013b) en alguna de las dos situaciones de evaluación sumativa: una prueba, por ejemplo, que se califica o bien la nota que se informa en un boletín.

Por eso las preguntas y cuestionamientos que se plantean en torno a la calificación no tienen que ver solamente con la nota en sí y las finalidades que hemos ido delineando, sino también con el proceso de calificación. Digamos que los problemas de uno son el reflejo de las inconsistencias del otro. Por otra parte, ha habido un desplazamiento en la finalidad de la escuela -en particular de la escuela secundaria- que hoy parece mucho más interesada en que todos puedan tener los aprendizajes esenciales (políticas llamadas de inclusión del más diverso tipo en nuestro país e incluso en Estados Unidos con “No Child left behind” del gobierno de Clinton). El surgimiento y evolución del sistema educativo y la irrupción de las notas como síntesis informativa del aprendizaje desde los comienzos del siglo XX (en sus diversas formas: números enteros, letras, porcentajes) tenían como central una función clasificatoria que facilitara establecer un ranking u orden de mérito que permitiera ubicar a los egresados en el mundo post-escolar del trabajo y la academia (O’Connor, 2010a). De modo que tenemos unas prácticas de calificación en una escuela media que tenía otra finalidad y que hoy en día, más allá de los itinerarios políticos

mencionados, tiene serias dificultades en perfilar claramente su sentido (Onetto, F., 2012).

Es necesario describir las prácticas de calificación y su lógica de producción en el concierto de estos sentidos que buscan legitimidad en el proceso de acceso masivo al nivel secundario que pegó un salto cuantitativo en nuestro país desde fines de los 90 con la consiguiente conmoción de un nivel educativo que estaba pensado para las élites o los becarios y que, en su crecimiento algunos pueden decir que ha caído en la errónea proliferación de lo mismo.

“La forma en que los docentes elaboran las notas dista mucho de ser evidente para todos los alumnos, y aún para sus padres. (...) funcionan como índices que anuncian que todo va bien, que la carrera académica sigue su curso, que el alumno trabaja con normalidad; o que hay que empezar a preocuparse, que está presente la amenaza de fracaso escolar” (Perrenoud, 2001: 144). Precisamente, el proceso de elaboración de las notas (la corrección de los exámenes, las escalas, sus grados y códigos de representación) han sido, como señalamos en el estado del arte, objeto de las primeras investigaciones en educación -por su centralidad en la experiencia educativa de los estudiantes- (Brookhart, Guskey, Bowers, McMillan, Smith, Smith, Stevens y Welsh, 2016) abordando desde las primeras décadas del siglo XX distintos aspectos relativos a dos grandes atributos que provienen en su definición teórica de la psicometría: la validez y la confiabilidad. Este último, comienza a investigarse con experiencias de multicorrección en Estados Unidos (Starch y Elliot, 1912; 1913a; 1913b) y en Francia (Laugier y Weinberg, 1927; 1930) dando lugar en este país a una nueva disciplina: la docimología (Martin, J., 2002) . En este caso son estudios dedicados a indagar la confiabilidad de la corrección y calificación de producciones y de exámenes nacionales (en el caso francés, el famoso Baccalauréat).

¿Cómo se elaboran las notas -especialmente las de un período o trimestre-? ¿De dónde surgen? ¿Qué información se tiene en cuenta? ¿Quién o quiénes intervienen en dicho proceso?

“Las notas de tareas individuales deberían reflejar los logros demostrados en el trabajo. Las notas de un boletín deberían reflejar los logros demostrados en un conjunto de trabajos para el informe de ese período” (Brookhart, 2017). Claro y sencillo, pero ¿cómo se realizan?

Por un lado, podemos describir los pasos que, más allá de ciertas diferencias metodológicas que tienen que ver con la validez de la calificación respecto de su uso e interpretación (Messick, 1989b; Brookhart, 1993; Sun y Cheng, 2014), son bastante comunes a todo proceso de calificación trimestral:

1. El docente enseña (con distintas estrategias didácticas) los contenidos del diseño curricular indicados para su asignatura y que ha organizado y secuenciado en su planificación.
2. Realiza algunas ejercitaciones y producciones que permitan ponerlos en práctica.
3. Califica todos o algunas de esas actividades que presentan los alumnos en forma individual o grupal durante el período trimestral.
4. Dichas calificaciones más alguna comprobación escrita u oral de eventual mayor peso en el trimestre que pretenda tener una función integradora, las integra en un promedio aritmético trimestral que vuelca en el boletín para que los padres tengan información acerca del desempeño del estudiante.
5. Hoy muchas jurisdicciones en nuestro país indican en sus regímenes académicos que los padres deben ser informados fehacientemente del proceso de aprendizaje y logros del estudiante antes del cierre del trimestre. Esto derivó en lo que se llama informe parcial que suele comunicar por escrito algunas de las calificaciones que se promediarán luego para conformar la calificación trimestral.

En los sistemas educativos de la Ciudad de Buenos Aires y de la Provincia de Buenos Aires cuyas prácticas de calificación indagamos en la presente investigación la acreditación final de la asignatura se realiza por el promedio de los tres trimestres entre sí. Deben arrojar un promedio mínimo de 6 (seis) puntos -y no menos de 6(seis) puntos en el tercer trimestre, en el caso de la Ciudad de Buenos Aires -por ahora RM 94/92 con sus modificatorias en Dirección de Programación educativa (2004)- y un mínimo de 7(siete) puntos de promedio -con no menos de 4 (cuatro) en el tercer trimestre) para acreditar la asignatura en la Provincia de Buenos Aires según su Régimen Académico (Dirección de Cultura y Educación de la Provincia de Buenos Aires, 2011).

En nuestro país, en la escuela secundaria o media siempre la acreditación es por asignatura, estableciéndose solamente que no puede adeudar más de 2 o 3 (según la jurisdicción) para ser promovido al curso superior una vez agotadas las distintas instancias de aprendizaje y evaluación del ciclo escolar con las complementarias que ofrecen en diferentes formatos en los meses de diciembre y febrero respectivamente.

En otros países, “combinado con las apreciaciones sintéticas para el conjunto de las disciplinas, se realiza un balance que contribuye a una decisión al finalizar el año escolar” (Perrenoud, 2008: 85), que derivan en consecuencias de cambio de curso, promoción al curso superior, repitencia, obtención o no de un certificado, etc. También nos encontraremos diferencias en los signos que representan los logros: notas, letras, porcentajes, conceptos; en las escalas: ordinales, intervalares, conceptuales; en las notas de corte o aprobación; en la ponderación de las distintas instancias de evaluación para la construcción de la nota trimestral. Pero lo esencial de la dinámica de la calificación tiene muy pocas diferencias en todos los sistemas escolares según las investigaciones e informaciones que hemos relevado.

Podemos caracterizar a las prácticas de calificación, según los métodos que emplean para elaborar la nota:

a. Calificación normativa.

Pone el foco en los resultados que permitan comparar los estudiantes de un curso entre sí y, eventualmente en grupos mayores: la escuela, el distrito, el sistema educativo como tal. Cuando prima la función clasificatoria de la evaluación, las notas van a cumplir, necesariamente, un rol fundamental en tanto que permitan establecer jerarquías. Recordemos que se llama normativa “en el sentido que fabrica una distribución normal o curva de Gauss” (Perrenoud, 2008: 86) y comparativa porque los resultados de unos se definen en relación con los de otros. Parece algo ajeno a las prácticas actuales de calificación, sin embargo, Marzano nos advierte que “aunque sean pocos o ninguno los esquemas de calificación actuales que utilizan un “enfoque normativo”, pueden encontrarse vestigios en las prácticas que elaboran un ranking de la clase y califican según la curva” (2010: 16).

b. Calificación referida a criterios.

Ya en la década del 30, Crooks comienza a plantear la necesidad de definir estándares absolutos de desempeño como base para la calificación, lo que en 1963 fue presentado como calificación “referida a criterios” en el señero trabajo de Glaser (Brookhart, 2009: 20). Uno de los métodos, dentro de este tipo de prácticas, se dio en llamar “*mastery grading*”: procuraba calificar al estudiante en función de su dominio de un nivel preestablecido. Se lo llamó también “*grading in time*”

ya que se mantenía el objetivo hasta que cada estudiante lo lograra en diferentes momentos del proceso de aprendizaje. Brookhart (2009) afirma que, a pesar de haber estado presente en la educación por casi 75 años, nunca tuvo gran difusión. Funcionaba- dice la autora- relativamente bien para conocimientos o habilidades que podían ser particionados en unidades, pero resultaba imposible de administrar para muchos estudiantes a través de una gran cantidad de objetivos de aprendizaje.

c. Calificación basada en estándares.

Podemos decir que es una variante de las prácticas referidas a criterio, va ganando en popularidad en el hemisferio norte. “La calificación según estándares de logro significa un cambio de pensar que las notas son lo que los estudiantes “ganan” a pensar en notas que muestren lo que los estudiantes aprenden” señala Brookhart (2017: 14) con claro optimismo.

Originalmente comparaba el desempeño del estudiante con un nivel preestablecido de desempeño, actualmente se suele referir a una lista de estándares de contenidos con sus respectivos niveles (*advanced, proficient, basic y below basic*) establecido por el estado o el distrito escolar (Brookhart, 2009). Precisamente la dificultad estriba en “establecer previamente una definición ajustada del rendimiento deseado”, nos advierte Camilloni (1998:149).

Se van difundiendo paulatinamente -sobre todo en nivel primario- boletines que dan cuenta del nivel que va adquiriendo el estudiante en distintos estándares por asignatura lo que ofrece un reporte mucho más rico de los desempeños que apunta a una evaluación formativa.

“A pesar del interés en este sistema -que muchos consideran el método más apropiado de calificación- todavía hay una escasa y pobre puesta en práctica junto a una considerable confusión acerca de las características que lo definen” (Marzano, 2010: 17).

Uno de sus aspectos interesantes es la incorporación de rúbricas, dándole a estas prácticas de evaluación y calificación una orientación formativa.

Figura 1. Esquema comparativo de prácticas de calificación.

Prácticas de calificación	Referencia	Tipo de calificación	Función
COMPARAR	Grupo clase- los demás compañeros	Normativa	Clasificatoria
ACREDITAR	Objetivos establecidos por los distintos niveles de política curricular.	Criteria	Certificativa y de rendición de cuentas
APRENDER	Su propio desempeño en relación a los estándares.	Basada en estándares	Formativa: brinda feedback para el docente y el estudiante

Fuente: Elaboración propia

Otro de los aspectos que van a configurar las prácticas de calificación son:

1. Qué producciones de los alumnos son calificadas y, por lo tanto, parte del proceso de evaluación sumativa que se concreta en el cierre trimestral. “Hasta hace poco, todo lo que estudiantes hacían, más allá de su objetivo era incluido en las notas” (O’Connor, K., 2010a: 3). Esto que O’Connor parece referir como ya superado, aparentemente sigue muy vigente en América Latina. Muchos profesores creen, aún hoy, que calificar más es evaluar mejor, que cuantas más notas tengan los alumnos en un trimestre, mejor se da cuenta de su proceso de aprendizaje.

“Conmigo tienen ocho, nueve o diez notas por período, porque yo les califico hasta los respiros”, según relata un docente (Ravela *et al*, 2017: 231).

Otros profesores afirman calificar todo y luego promediar con una nota de concepto que remite a aspectos no académicos del aprendizaje como el esfuerzo, la responsabilidad, la asistencia, la atención durante la clase, el cumplimiento en tiempo y forma con las tareas asignadas para hacer en la casa, el asistir con los materiales indicados para las actividades planificadas por el docente para el aula y un gran etcétera que depende de la asignatura, del nivel educativo y también de algunas cuestiones idiosincráticas .

Todo termina en una sola calificación que no permite saber a qué obedece, qué significa. Este rasgo que Brookhart llamó “hodgepodge grading” (Brookhart, S., 1991: 36) y que otros autores (Celman, S., 1998; Cizek, 1999; Hummel-Wiley,) han expresado con otras metáforas da cuenta de la dilución de la validez (no se sabe lo que mide e informa, por lo tanto, difícilmente podemos dar cuenta de que mida tal o cual cosa). Es todo un gran revoltijo que confunde y, precisamente por su ambigüedad, termina constituyéndose en un instrumento de disciplinamiento y motivación externa que sustituye el interés natural por aprender (Ravela *et al*, 2017).

2. Un fenómeno ampliamente extendido en América Latina y fuertemente arraigado en las normativas y en las prácticas de calificación de la escuela secundaria argentina tiene que ver con la “matematización de las calificaciones (...) que induce a los docentes a construir la calificación como una suma y ponderación de puntajes, perdiendo de vista los desempeños (...) genera una apariencia de objetividad, pero carece de significado sustantivo en términos de lo aprendido” (Ravela *et al*, 2017: 227). En la cultura del docente -y pareciera muy vigente en la sociedad argentina- hay un carácter sagrado del número que lleva a otorgarle una fantasía de precisión. Además de que el número responde a un sistema formal matemático de descripción de la realidad y, por tanto, uno posible entre otros, suponer que los aprendizajes se pueden representar con números lleva a cuestiones profundas de validez -que queremos abordar especialmente en esta investigación- ya que se dificulta asignar un significado sustantivo a una cifra determinada o incluso una diferencia en logro entre un 6.50 y un 6,80 -por ejemplo- (Ravela *et al*, 2017).

3. Los promedios

Además de la utilización de números para representar niveles de desempeño o aprendizajes, nuestro sistema educativo (en particular las jurisdicciones cuyos docentes consultamos en esta investigación) introduce la operación con estos números en función de:

- a. Generar un promedio a partir de la combinación de las notas numéricas de distintas producciones o desempeños para la asignación de una nota resumen de un período trimestral que se informa en el boletín de calificaciones.
- b. Y, a su vez, la operación de promediar entre las notas trimestrales -con distintas modalidades según jurisdicción- para establecer a partir del resultado la acreditación de la asignatura o que el estudiante debe continuar con nuevas instancias de aprendizaje y/o evaluación en períodos complementarios.

La nota trimestral, que es objeto de los escenarios de calificación que hemos ofrecido a los docentes para expresar sus eventuales prácticas y decisiones, tiene la peculiaridad de “combinar o mezclar” distintos aspectos y producciones del estudiante: aquellos que se llaman factores académicos -pruebas, quizzes, ensayos o trabajos prácticos escritos, exposiciones orales, presentaciones escritas o digitales, etc- que han sido objeto de evaluación sumativa y a los que se les ha asignado una nota y, también, pese a las recomendaciones que vienen haciendo hace tiempo los expertos en evaluación y calificación (Brookhart, 1991; 1993,2009; McMillan, 2001, 2019; O’Connor, 2009, 2010a,2013), toda una serie de factores no-académicos que generan un *merged judgement* (Brookhart, 2013) y que se traduce en el fenómeno que se describe con la metáfora conceptual de “hodgepodge grading” (Brookhart, 1991) que ya hemos mencionado. Como lo que se califica se promedia, no sólo es confuso qué, por qué y cómo se califican algunas producciones o conductas, sino que, además, al ser promediadas notas que parten de evidencias de aprendizaje totalmente diversas, el significado de esa nota es totalmente confuso. Algunos docentes apelan a calificar los que se llamó los tres tipos de contenidos: conceptuales, procedimentales y actitudinales, que tuvieron amplia difusión como modo de diseño curricular y planificación en las transformaciones educativas en nuestro país fruto de la Ley Federal de Educación

(1993) y, desde esa perspectiva, consideran que una mirada integral del aprendizaje debe calificar también aspectos actitudinales (esfuerzo, cumplimiento de tareas en tiempo y forma, actitud y participación en clase, etc.). Otros, tradicionalmente utilizan la nota de concepto como mecanismo regulador para aquellos estudiantes que están en una situación límite de aprobación. Ravela (2017) nos advierte que esa búsqueda de integralidad puede derivar tanto en notas sin significado concreto como en evaluaciones fuertemente disciplinadoras.

4. Estas notas de concepto, promedios que son fruto de mezclar distinto tipo de factores y eventuales redondeos o ajustes que realiza el profesor determinan aspectos intersubjetivos de la calificación que algunos autores llaman negociación o fabricación (Merle, 2007; Perrenoud, 2009).

Son estas modalidades las que favorecen los *arrangements* -arreglos, acuerdos- más variados, que tan profundamente analiza Merle:

“la nota no es solamente una medida de competencias. Es también una mediación entre los profesores y los estudiantes conocidos e identificados por sus nombres. Las prácticas de calificación permiten mostrar que, lejos de ser medidas automáticas del “nivel” de los estudiantes”, son, de modo variable, el producto de estrategias y de coerciones específicas (...) estos contextos escolares están en relación con procesos de fabricación de las notas fundados en diferentes tipos de arreglo” (Merle, 2007: 54).

Estos aspectos intersubjetivos de la calificación que algunos autores llaman fabricación y/o negociación. Como la nota se transforma en “el” objetivo de los estudiantes (Shepard, 2006)

Tanto Camilloni (1998) como Withney (2006) señalan que, más allá del sistema que se utilice, la calificación se convierte en un elemento crucial del sistema educativo. No es un dispositivo anecdótico ni ingenuo. La elección del sistema, su eventual combinación con otros e interpretación estará directamente relacionada con el estilo de enseñanza-aprendizaje que tenga el profesor.

5. Los grados y las escalas de calificación.

Los grados -señala Camilloni (1998)- son el número de categorías establecidas en términos descriptivos a las que se le atribuye de antemano valores numéricos convencionales. En el caso de la normativa de evaluación de la Ciudad de Buenos Aires se trata de 11 grados, del 0 al 10, siendo 6 correspondientes a la desaprobación y 5 para la aprobación. Por su parte, en este aspecto concreto de la normativa, la provincia de Buenos Aires determina 11 grados, del 0 a 10, siendo 7 correspondientes a la desaprobación y 4 para la aprobación. Hay, dice

Camilloni, una “propuesta de reducción del número de grados de las escalas de calificación (...) fruto de las corrientes anticuantitativistas en educación (...) que las consideran un exceso en la diferenciación cuantitativa aplicada al análisis de las realizaciones de los alumnos” (1998: 138). Por otra parte, Uruguay tiene una escala de 0 a 12, Francia de 0 a 20 y, siempre refiriéndonos a la escuela secundaria, Inglaterra y Estados Unidos “detrás” de las calificaciones en letras que suele utilizar (A-F en sentido descendente con algunas variantes) conservan una escala de 0 a 100 con cada puntaje bastante formalizado en su asignación a cada letra.

Significado y validez de las calificaciones

Como hemos venido señalando, los problemas de las prácticas de calificación que atraviesan las diferentes tensiones de los interesados *-stakeholders-* que establecen distintas demandas y finalidades a las notas, la forma de elaboración que incluye algunos mecanismos de supuesta objetividad matemática con serias falencias estadísticas que mezclan apreciaciones sobre el desempeño del estudiante de distinto orden (académico y actitudinal), una construcción que guarda cierta confusión y arbitrariedad por parte del docente, sumado a un sentido, interpretación y uso diverso nos están marcando como central una cuestión de validez.

Señala Angulo Rasco (en Celman, S, 1998: 52) que, “creer que las calificaciones son los únicos elementos necesarios para formular un juicio de valor, supone desconocer u ocultar la participación, normalmente inconsciente, de otros elementos como los valores que sustentamos, nuestras concepciones educativas y docentes”. De allí el concepto de “signo opaco” con el que Susana Celman describe el uso que se hace de las notas.

En un primer informe sobre el tema Susan Brookhart plantea -y luego continuará desarrollándolo en toda su carrera de investigadora y capacitadora cuyos materiales y enfoque teórico seguimos- que “esta sensación de discrepancia e incomodidad que experimentan los docentes entre las prácticas de calificación recomendadas y las reales son síntomas de un problema de validez” (1991: 35).

La psicometría y las evaluaciones masivas estandarizadas han planteado desde sus comienzos el tema de la validez unido a la confiabilidad. La aplicación de un enfoque psicométrico tradicional de validez a la evaluación en el aula de clase ha fallado,

Las múltiples características específicas de la evaluación en el aula tienden a limitar la capacidad del enfoque de la psicometría tradicional para recoger evidencia que ratifique o refute los argumentos de validez (Bonner, 2013).

Señala Camilloni que “el concepto de validez, en términos generales se refiere a la capacidad de un instrumento de evaluación para medir lo que se pretende evaluar con él” (2010: 26) y aclara que nunca es absoluta, sino contextual pues -además de la confiabilidad- su utilidad y practicidad van a configurar la pertinencia de un programa de evaluación (Camilloni, 1998).

Uno de los autores que más han trabajado la validez y cuyos trabajos han marcado en lo sucesivo la discusión sobre el tema, es Samuel Messick. Si bien plantea distintos tipos de validez: de contenido, criterial, de construcción, señala que “no es adecuado, diferenciar así estas formas de validez porque abren la posibilidad de que los instrumentos posean uno o dos tipos de validez y no todos” (Camilloni, 2010: 27). Es un concepto que no ha tenido una definición total y plenamente acordada entre los especialistas como tampoco su importancia es siempre estimada en el mismo grado (Camilloni, 1998, 2010).

De entre los distintos enfoques, nos quedaremos precisamente con su definición:

La validez es un juicio evaluativo general del grado en que la evidencia empírica y los fundamentos teóricos apoyan la exactitud y propiedad de las interpretaciones y acciones basadas en los puntajes de una prueba u otras formas de evaluación (Messick, 1989a: 54).

No es una propiedad de la prueba o de la evaluación en sí mismos, sino de significado de los puntajes de las pruebas (Messick, 1995).

Esta definición y enfoque teórico de Samuel Messick es el que va a tomar Brookhart (1993) para analizar el problema de validez de las calificaciones que asignan los profesores de educación secundaria y superior ya que considera que las prácticas de calificación de los profesores reflejan una preocupación por las consecuencias que acarrearán las notas, muchas veces a expensas de la consideración de su interpretabilidad. “El significado de una nota y los valores implicados en su uso e interpretación están interrelacionados con el concepto de validez” (Messick, 1989a) y son cruciales para la descripción de las prácticas de calificación” (Brookhart, 1993: 123).

“Si Ud. quiere utilizar información para comunicarse con los estudiantes y los padres, para tomar decisiones sobre su enseñanza, para darle feedback a los alumnos que utilicen en sus trabajos sucesivos o para cualquier otro propósito, necesita estar seguro de que la información es la correcta” (Brookhart, 2009: 9), la mejor, la más precisa y clara.

Es precisamente la validez la que dará cuenta de la significatividad y propiedad de una información (de una nota).

También en esta investigación vamos a utilizar el enfoque de Messick, sin desconocer que el debate continúa y va evolucionando e incorporando otras miradas de la validez, algunas más orientadas a la evaluación estandarizada a gran escala y otras que se plantean como más apropiadas para la evaluación en el aula. En ese sentido plantea Robert Marzano que

La perspectiva basada en argumentos está perfectamente ajustada al docente del aula y ellos son los individuos más capacitados para generar la red de sentidos e inferencias que plantea Kane (1992, 2001, 2009) que permitan llegar a conclusiones y decisiones a partir de los puntajes generados por la evaluación en el aula (Marzano, 2018: 21).

Sin embargo, podemos encontrar una serie de conclusiones compartidas por los distintos autores, a saber:

- ➔ La validez no es la propiedad de una prueba o de un resultado (nota) o de una nota combinada (promedio trimestral) sino que se refiere al uso que se le da para una finalidad específica.
- ➔ Para evaluar su utilidad y su adecuación para un propósito particular se requieren múltiples fuentes de evidencia.
- ➔ Un test o instrumento no tiene una validez general, sino que debe ofrecerse evidencia suficiente que es adecuado para ese fin concreto que se procura.
- ➔ La evaluación de la validez de una prueba de evaluación no constituye algo estático, sino que es un proceso continuo. (Camilloni, 2010).

La matriz progresiva de validez en Messick (1989 y 1993).

La propuesta teórica de Messick (1989 y otros) reencuadra el concepto de validez y le provee de una mayor riqueza, complejidad e inserción contextual. De este modo da cuenta de aquellos elementos que parecían quedar “por fuera” de la validez incluyendo todas sus dimensiones en el encuadre marco de la validez de construcción a la que presenta desde sus dos fuentes de validación: la justificación y la función intencional. Queda configurada así la que él dio en llamar “matriz progresiva de la validez” (Messick, 1989b: 10) porque va creciendo y complejizándose en la medida que cruza las dos intenciones de interpretación y uso con la justificación que deriva de las evidencias que permiten validar un test, una nota y/o un resultado como así también las implicancias

axiológicas (valores, supuestos, creencias) y las consecuencias sociales que, según el contexto, pueden derivar de su uso dentro o fuera del sistema (el escolar en este caso).

Figura 1: Matriz progresiva de validez.

Fuente de justificación	Función intencional	
	Interpretación	Uso
Evidencias	VC (validez de construcción)	VC + RU (relevancia y utilidad)
Consecuencias	VC + IV (implicancias de valor - axiológicas)	VC + RU + IV + CS (consecuencias sociales)

Fuente: Adaptado y traducido de Messick, 1989b.

Este mismo modelo utilizaremos -como lo hacen Brookhart (1993) y Sun y Cheng (2014) - para analizar la validez de las prácticas de calificación de los docentes de escuelas medias de Ciudad de Buenos Aires y Provincia de Buenos Aires a partir de los comentarios que hacen al explicar las decisiones que han tomado al asignar una nota a un estudiante.

Cada uno de los cuadrantes presenta las distintas facetas de validez que surgen de la matriz.

La validez de construcción (VC) refiere al significado de la nota en sí misma a partir de los criterios de construcción: cómo es el mecanismo del promedio que establece el docente o la normativa, qué incluye o excluye de las evaluaciones sumativas que ha hecho el estudiante, qué otro tipo de factores no académicos va a incorporar a la nota trimestral (esfuerzo, participación, compromiso, tiempo y forma de entrega de trabajos), etc.

La relevancia y utilidad (RU) refleja a partir de qué evidencias de desempeño del estudiante y/o del docente se ha elaborado la nota trimestral del alumno. Apunta a señalar qué significa esa nota para ese estudiante en esa circunstancia particular.

Las implicancias de valor (IV) procuran explicitar los valores subyacentes a la decisión de calificación que el docente ha realizado (si es justo o no, equitativo, si dicho aspecto lo relaciona también con el resto del grupo, si el docente ha tenido alguna actitud compasiva ante una dificultad conocida del estudiante, etc.)

Las consecuencias sociales (CS) refieren la validación de la calificación a las consecuencias que pueda tener la nota dentro del ámbito escolar (cómo continúa con la materia, en otras materias, en el colegio, en su escolarización) o fuera de él: en la aplicación e ingreso a estudios superiores, en el curriculum vitae que deberá presentar para acceder a un puesto de trabajo, etc.



Universidad de
San Andrés

CAPITULO 3

HISTORIA DE LAS CALIFICACIONES

1. ¿Para qué una historia de las calificaciones?

Al abordar la descripción de una práctica social en el marco de una investigación, se hace necesario encuadrar no sólo su contexto actual de producción, sino también establecer el origen de dicha práctica, su evolución y las distintas perspectivas desde las que ha sido abordada por la teoría y/o la investigación empírica.

¿Por qué es importante tratar de rastrear el origen de las calificaciones, sus normativas y sus prácticas?

Porque, en tanto práctica social, su origen y evolución no ha tenido una dinámica lineal, clara y definida. Por la importancia que tienen las prácticas de calificación frente a todo intento que se pudiera llevar adelante de transformación de las estrategias de enseñanza, aprendizaje y evaluación, ya que “suelen operar de manera muy efectiva, en reversa, como determinantes de todas las prácticas pedagógicas” (Camilloni, 1998:133).

Señala Perrenoud que el peso del pasado en las estructuras, en los textos y sobre todo en las cabezas es demasiado fuerte frente a cualquier intento de transformación en el mundo escolar ya que, “el siglo XX ha demostrado la fuerza de inercia del sistema, más allá de los discursos reformistas. En tanto muchos pedagogos han creído enjuiciar definitivamente las notas, éstas están todavía allí, y bien arraigadas en numerosos sistemas escolares” (Perrenoud, 2008: 8).

Todos percibimos la calificación a partir de nuestra propia experiencia y nuestra reflexión está limitada por los cambios que hemos experimentado como alumnos y como docentes (Withney *et al*, 2006). Esto da una “ilusión de transparencia” (Bourdieu, 2002) -tal como lo mencionábamos en el capítulo teórico- frente a aquellas prácticas que son muy cercanas a la cotidianeidad (como se trata de la calificación en los profesores). Lleva a pensar que siempre se hizo de este modo porque es así, inherente a la práctica docente (Carlino, 1999) y no puede ser hecho de otra manera. “Toda persona que evalúa tiene incorporado – en forma explícita o implícita- un modelo o un ideal de evaluación” (Gvirtz, 2008: 262). De ahí que se hace necesario explicitar el propio y rastrear sus orígenes en la “historia de calificado” y en la construcción histórica que se plasmó en diversas normativas que procuraron articular las prácticas de calificación.

El docente experimenta la presión de la corrección, la calificación y los informes como la tarea más abrumadora y compleja según reflejan encuestas desde mediados de la década del 70. Por eso Cardinet propone “estudiar el origen de esta dificultad, es decir las razones que llevaron a consagrar un tiempo excesivo al control” (1989: 41).

Una aproximación histórica a las prácticas de evaluación “puede constituirse en el requisito previo de un camino que nos lleve a identificar sus funciones y el modo en que operan” (Barbier, 1983: 47)

Podemos afirmar, con Florencia Carlino, que “los atributos que, a simple vista, aparecían intrínsecamente ligados a la evaluación, lejos de ser parte de su esencia, son las características que asumen las prácticas concretas –social e históricamente situadas- de la evaluación en nuestra sociedad” (1999: 20). “Las calificaciones han sido parte de la escolaridad por generaciones. Se han vuelto tan generalizadas y predecibles que los docentes, alumnos y autoridades las ven como esenciales e inevitables” (Malehorn, 1994: 323). Una mirada retrospectiva que permita sondear en las fuentes originales de la calificación puede arrojar algunas respuestas que nos permitan entender mejor las prácticas actuales (Durm, M., 1993).

Un campo poco y superficialmente transitado

Los autores que han comenzado a intentar sistematizar una historia de la evaluación y, muy especialmente, la historia de la calificación (Cureton, 1971; Barbier, 1983; Durm, 1993; Guskey, 1994; Moll, 1998; Maulini, 2003; Guigou, 2006; Merle, 2007) coinciden en señalar la escasez de la investigación acerca de la génesis y evolución de las calificaciones. “La propuesta de una perspectiva histórica de las modalidades de calificación que han tenido lugar dentro de nuestro sistema educativo ofrece un enfoque estimulante pero arduo, en tanto las investigaciones sobre este tema son escasas” (Chambat, 2010: 1).

Por otra parte, Guigou sostiene que casi todos los trabajos que abordan la historia de la evaluación carecen de una crítica interna, ya que

...casi nunca cuestionan la legitimidad política de las prácticas de evaluación (...) sin desprenderse del evolucionismo y del progresismo según los cuales habríamos pasado de las pruebas y mediciones –reductivos e indiferenciados- en los orígenes de la docimología, a los dispositivos ampliados, autonomizantes, diferenciados y reguladores de la evaluación formativa y de la evaluación-investigación de hoy (Guigou, 2006: 1).

Esta acusación de presupuesto triunfalista (Guigou, 2006) se presentaría en los trabajos que hemos leído del ámbito anglófono y que se limitan a describir la evolución

de la evaluación y las calificaciones (Cureton,1971; Durm, M.,1993; Gipps, 1994; Guskey, 1994;O'Connor, 2010; Shepard, 2006) señalando caminos para su optimización, pero sin una mirada crítica que cuestione la evaluación misma y su articulación con las prácticas de clasificación, ejercicio del poder, la autoridad y la selección social.

Otra vertiente más crítica parece presentarse en el análisis del propio Guigou (2006), Merle (2007) y Bugnard (2004) quienes presentan la historia de las calificaciones en el marco de las transformaciones del capitalismo que generaron la escuela moderna y su consiguiente necesidad de clasificación, como así también en el marco de la evolución de las tecnologías del castigo y disciplinamiento de los cuerpos –con una mirada claramente foucaultiana-.

Algunos autores como Barbier (1983) y Perrenoud (2001,2008) han profundizado en este enfoque, influenciados por Bourdieu y Foucault, para dar cuenta de la complejidad de la evaluación de los aprendizajes a través de los conceptos de fabricación de la excelencia, el éxito y fracaso escolar, la selección, el reconocimiento, la negación de las desigualdades y las clasificaciones. De Ketele (1993) procura un itinerario que podríamos llamar de síntesis epistemológica, al presentar a los autores más influyentes, desde la aparición, desarrollo, evolución y vigencia de sus postulados en tanto vertebradores de diversos y móviles paradigmas.

Entre los autores que, en español, ofrecen una mirada crítica, se destaca el provocador trabajo sobre la historia del examen de Ángel Díaz Barriga (1994) quien cuestiona la misma posibilidad de medir los aprendizajes, la naturalización del examen como corolario de toda acción educativa y su fetichización armónica con el concepto de calidad educativa.

En nuestro medio, Florencia Carlino, al ofrecer un panorama histórico sobre la evaluación, incorpora su análisis como práctica social, con el propósito declarado de “proveer los elementos contextuales de cada época, reseñar procesos, presentar documentos y protagonistas claves a fin de estimular al lector a la reflexión acerca de las continuidades y las rupturas entre el presente y el pasado” (1999: 29).

Llevar adelante ese enfoque de las calificaciones como producto y dispositivo de control social que nace con la masificación de los sistemas educativos del capitalismo de mediados del siglo XVIII, tanto Perrenoud (2001; 2008) como Maulini (1996; 2003) y procuran, en un sano realismo, aceptar que la evaluación y la nota ya están ahí y que –sin olvidar su origen- se pueden mejorar las prácticas de evaluación y calificación. Digamos

que dan cuenta de la complejidad de la construcción del campo al plantear un permanente diálogo entre la didáctica, la pedagogía y la reflexión sociohistórica y política.

Hay muy pocos trabajos dedicados específicamente a la historia de las calificaciones o notas, si bien no podemos dejar de tener una mirada sistémica en la que éstas están mutuamente determinadas con los marcos teórico-metodológicos que se sustentan en el proceso de enseñanza, aprendizaje y evaluación.

2. Génesis y desarrollo de la evaluación y las calificaciones

Antecedentes remotos

Varios autores hablan de una cierta prehistoria de la evaluación en la burocracia china (mandarinato) para elegir miembros de las castas inferiores (Díaz Barriga, 1994; Cureton, 1971). Allí construyeron un sistema de examinación que, según Stufflebeam y Shinkfield, permanece vigente hasta la actualidad para seleccionar los mejores candidatos a ocupar los cargos públicos en el gobierno chino (en Carlino, 1999).

Esbozos de evaluaciones siempre hubo en la historia de la humanidad: pruebas de valor, de pureza, de idoneidad. Algunas tenían como resultado la muerte, la posibilidad de seguir viviendo (gladiador) o bien el paso a una élite (cazador, brujo, mandarín) o status social (ritos de iniciación para la inserción como adulto en la tribu).

La Edad Media ve aparecer las universidades y las escuelas. En las primeras, había algunas formas de evaluación para formar parte del claustro docente o tener una licencia de profesor que, particularmente la de la Universidad de París, permitía ejercer en toda Europa. Por otra parte, existía cierta exposición pública del saber a la que no eran sometidos los alumnos hasta tanto sus profesores hubiesen establecido su adecuada preparación (Durkheim, E. citado por Díaz Barriga, 1994) Era más una instancia de lucimiento y de publicitación de la categoría, el saber y prestigio de un determinado discípulo, que un examen moderno en el que se pusiera a prueba qué sabe un estudiante.

En esos tiempos, los más pobres no están escolarizados. Algunos pueden recibir una enseñanza brindada por la Iglesia en forma muy rudimentaria articulada en torno al catecismo. Sin embargo “junto a estas grandes instituciones (universidad, escuela) hay un tejido informal de “pequeñas escuelas” que permiten una relativa democratización del saber” (Maulini, 1996: 2). Allí, el maestro era retribuido en especias según lo que aprendía el alumno (Chambat, 2010). Por tanto, podríamos decir que eran las familias las que

tenían la responsabilidad de la evaluación y el maestro estaba más que interesado en que el alumno aprenda, a riesgo de perder un cliente (Maulini,1996).

El estudio representaba, en la Edad Media, un modo de vida más que una preparación a la vida. La responsabilidad de decidir si había algún interés espiritual o material en estudiar o continuar haciéndolo residía en el alumno y su familia. Cierta relación, dice Perrenoud (lo cita Maulini,1996), con un consumo escolar que se expresa en la selección de escuelas que hoy realizan las familias. Comenio, en su *Didáctica Magna* (1657) incluye al examen como parte del método, pero no decide ni la promoción ni una calificación del estudiante, nos recuerda Díaz Barriga (1994).

La bisagra: Los Jesuitas

Desde mediados del siglo XVI la Compañía de Jesús abre colegios por toda Europa, siendo su tarea educativa uno de los principales brazos de la Iglesia contra la Reforma Protestante.

Primero abrieron los colegios a novicios (futuros religiosos jesuitas) y luego paulatinamente a la burguesía. Era una educación gratuita y elitista, en tanto procuraba conformar una juventud católica, instruida y disciplinada que lidere los distintos estados europeos (Maulini,1996).

Ya no se trata del privilegio de cuna, sino que una nueva jerarquía de mérito será la que clasifique y seleccione. Su código de enseñanza, aún vigente con reformas, era la *Ratio Studiorum* (publicada en 1599). Este auténtico curriculum organiza a la clase en dos ejércitos (romanos y cartagineses) con todas las jerarquías individuales. Establece una maquinaria de recitaciones, composiciones, disputas, concursos, premios, justas de elocuencia, deberes escritos, revisiones cotidianas, mensuales, trimestrales y anuales. Ambos grupos compiten entre sí y cada alumno tiene su directo opuesto en el otro equipo. Las jerarquías y dignidades se renuevan cada mes en los resultados de esta competencia.

Los investigadores europeos y latinoamericanos señalan a los colegios jesuitas como precursores de diversos rasgos que luego asumirá la escuela moderna: la ritualización, el control y el disciplinamiento, la emulación y la competencia relacionando al proceso de aprendizaje con el premio como estímulo externo al propio saber.

Es precisamente el Colegio Jesuita de Coimbra (Portugal) que inventa en 1558 la primera distribución de premios, extendiéndose rápidamente la idea a los demás colegios de la Orden en Europa (Chambat, 2010: 2).

Veamos el desplazamiento de varios elementos del aprendizaje:

Figura 3: De la Edad Media a los Colegios Jesuitas

	Edad Media	Colegios Jesuitas
Maestro	Guía- Tutor -Eje en el saber	Vigilante –Preceptor -Eje en la virtud
Alumno	Autonomía - Aprende	Puerilizado - Emula
Grupos- Aulas	Heterogéneos	Graduación por edad
Examen	Pura formalidad	Central –Competitivo- Individualista
Éxito	Mediato: el propio saber	Inmediato e individual: premio

Fuente: Elaboración propia

Había que establecer una clasificación que partiera de la contabilización de los errores en las composiciones escritas, los que eran informados a los padres indicando el puesto del alumno en un ranking. A fin de año los estudiantes son clasificados en 6 categorías que van de los mejores a aquellos que deben repetir: los excelentes, los muy buenos, los buenos, los satisfactorios y los mediocres. Los dudosos (dubii) son admitidos a la clase siguiente “a prueba”. Si hay problema vuelven al curso anterior. Se solicita, finalmente, que los padres retiren a los -así llamados- limítrofes y burros. (Maulini,1996; Bugnard, 2004).

“Pronto las categorías son traducidas en números, del 1 al 6. Una vez que la nota numérica es introducida, el resultado punitivo despliega una economía rentable del sistema escolar, un sistema que rápidamente se extendió a toda la escolaridad obligatoria, en el siglo XIX, con diversos tipos de escalas” (Bugnard, 2004: 93).

De los registros de una escuela de Londres surge que, hacia 1650, los cursos se examinaban entre sí en equipos, y el “equipo ganador ocupaba el asiento más alto de la escuela” (Cureton,1971: 2). Un antiguo boletín del Pensionado del Colegio Real de Cahors, describe al alumno con una palabra en cada asignatura: excelente, muy tímido, trabaja, etc. e incluye el puesto que ocupa en cada mes sobre un total de 52 alumnos.

Sin la influencia de la “constante macabra” (Antibi, 2012) o la distribución de la curva normal (Cureton, 1971) ya podemos ver que la evaluación va dando lugar a la clasificación y la calificación como un dispositivo de selección social que va teniendo distintas formas de expresión.

Barbier, en un profundo análisis, sostiene que en las formas sociales pre-capitalistas “las prácticas de evaluación de las personas en formación tienden a aparecer

cuando se reúnen dos condiciones: una relativa autonomización del factor “trabajo” y una diferenciación de este factor en el ejercicio de funciones dependientes y jerarquizadas” (1983: 48). Cuando se dan estas condiciones arriba mencionadas, en los colegios, escuelas y universidades se establece un mecanismo de reconocimiento social de las capacidades del individuo.

Es lo que el propio Barbier llama “generalización de las prácticas de evaluación en las sociedades precapitalistas”, las que interrelaciona con una matriz muy clara:

Figura 4: Condiciones históricas de desarrollo de las prácticas de evaluación

Campo de aparición de los fenómenos	Fenómeno relevante de los procesos de evaluación de los individuos	Fenómenos relativos a los sistemas sociales en cuyos cuadros se realizan estas evaluaciones	Ideología dominante en estos sistemas o subsistemas
<i>Producción de los medios de existencia</i>	Determinación de los salarios y de jerarquías (reconocimiento del valor de la fuerza de trabajo)	Organización de un mercado de empleo a partir de una población “activa” diferenciada.	Ideología del mérito social (el mejor en el mejor lugar)
<i>Ejercicio del trabajo</i>	Clasificación profesional a partir de una jerarquía de puestos y calificaciones	Organización de sistemas de trabajo alrededor de una división de tareas y de una dependencia funcional de actividades	Ideología de la competencia profesional

<i>Actividades de formación</i>	Exámenes, diplomas, certificados, etc	Desarrollo de un sistema educativo.	Ideología de la igualdad de oportunidades escolares
<i>Actividades pedagógicas en sentido estricto</i>	Calificación	La clase como grupo estable y rígido	Ideología de la emulación pedagógica

Fuente: Barbier, 1983: 49

Barbier (1983) nos describe la evolución de un sistema educativo que funciona como una institución autónoma, completa y especializada en la función de producción de capacidades, organización en grados y niveles. Hasta aquí esta prehistoria de la evaluación y la calificación.

La invención de la nota

En los mismos albores de la masificación escolar, ya los Hermanos de las Escuelas Cristianas tenían su propio sistema de castigos (por ejemplo, los *pensum* o tareas de penitencia), y privilegios. Todos estos se reunían en un sistema de puntos informado a las familias. La clasificación de los alumnos en grupos jerarquizados va a culminar en la calificación de cada una de las cohortes. “Sea que se presenten en forma de anotaciones concretas o escrituras simples, letras o números, se sitúen en una escala graduada de 0 a 20 o de 0 a 6, estas notas surgirán de la selección imaginada por los jesuitas y sus contemporáneos” (Maulini, 1996: 6).

“La calificación escolar es un acto que se materializa en el siglo XIX” (Díaz Barriga, 1994: 175). Esta contundente afirmación que ratifica Florencia Carlino, es compartida por todos los investigadores. Así, por ejemplo, O’ Connor destaca que “si bien la evaluación y el reporte de los logros del estudiante han sido parte de la educación por siglos, sin embargo, el uso de las notas tiene un desarrollo mucho más reciente” (O’ Connor, K, 2010: 38), lo que comparten por ejemplo Chambat: “Muchos colegas se sorprenderán probablemente al enterarse que la calificación es una idea nueva. Lo que

hoy parece una evidencia admitida por todos, es en realidad el resultado de un largo proceso” (2010:1) o Thomas Guskey (1994) al describirlo como una práctica prácticamente desconocida en los Estados Unidos hasta 1850.

La información brindada a los padres y a los propios estudiantes acerca de su progreso era en general de tipo oral (Pollio, H. 2004: 1) en un “formato narrativo y frecuentemente sólo enunciaba los conceptos y habilidades que un estudiante había adquirido” (O’ Connor, 2010: 38).

¿Será imposible con las tecnologías que hoy disponemos ofrecer una descripción de los logros en formatos nuevos, integrales que completen la nota numérica?

Las primeras calificaciones sistemáticas aparecen sobre fines del siglo XVIII, utilizando distinto tipo de rangos o sistemas, a partir de la primera experiencia en la Universidad de Yale que utiliza una serie de adjetivos descriptivos hacia 1785. El decano Stiles, en el pie de página de su diario, anota las calificaciones de 58 estudiantes que habían rendido examen: *Optimi; Second Optimi; Inferiores (Boni); Pejores*. Dice Durm que, “con toda probabilidad éstas han sido las primeras calificaciones académicas otorgadas en los Estados Unidos” (1993: 2). Luego incorporó la propia Yale una escala de 4 puntos y otras universidades como Harvard (hacia 1810), Virginia y Georgia (1891) y Michigan desde comienzos del siglo XIX.

Lo llamativo es que, desde el comienzo de su implementación, surgen las siguientes dificultades:

- a) Se intentan corregir permanentemente las distintas escalas, rangos y sistemas. Así, por ejemplo, entre 1800 y 1832, la Universidad de Yale no cambió menos de 4 veces de sistema: desde el descriptivo a la escala de 4 puntos, otro basado en 9, para finalmente volver a 4. La Universidad de Harvard comienza agrupando a los estudiantes en 6 divisiones en 1877 (sobre una base de 100 puntos), aparecen las primeras “letras” en 1883, 5 o 6 niveles, aprobado-desaprobado hacia 1895, para finalmente utilizar las letras que luego se consagrarían en todo el sistema educativo americano. (Durm, 1993; Cureton, 1971)
- b) Hay quejas de lo que hoy llaman los autores *grade inflation* (inflación de notas), denunciando Henry Ruffner en 1890, que hay una “continua tendencia a calificar demasiado alto a estudiantes inferiores” (Cureton, 1971: 2).

A partir de las leyes de asistencia obligatoria en el nivel primario a fines del siglo XIX y comienzos del XX, aumentó muchísimo el número de estudiantes que ingresaban a la escuela secundaria. “Así, entre 1870 y 1910, el número de secundarias públicas en los Estados Unidos aumentó de 500 a 10.000” (Pollio, 2004:1), con un crecimiento similar en Francia, Alemania, Holanda, Bélgica, Suiza. Las escuelas de aula única comenzaron a desaparecer y ser reemplazadas por escuelas más grandes organizadas por edad de los estudiantes. Lógica y necesariamente la evaluación evolucionó de acuerdo con estas tendencias.

A fines del siglo XIX, se dan dos hechos significativos en la historia de la calificación:

- En Francia, Jules Ferry va a promulgar un decreto el 5 de junio de 1890 que estipula que cada composición tendrá una nota de 0 a 20. Este sistema perdura masivamente hasta hoy en la escuela francesa.

- En Estados Unidos, en el College Mount Holyoke (1897) se establece el primer sistema de calificación con letras, si bien era una combinación de adjetivos descriptivos, letras y porcentajes (Durm,1993).

Entre 1911 y 1960 los sistemas educativos fueron experimentando con varios códigos de letras y números para informar el rendimiento de los alumnos. Si bien fue muy popular el sistema de escala centesimal en la primera parte del siglo XX, hacia la década del 40 ya la mayoría de las escuelas utilizaban el sistema de letras en Estados Unidos. “En la década del 60, la lucha por humanizar las escuelas empujó a algunas instituciones a desplazarse hacia la opción de aprobado/desaprobado en las pruebas escritas. Sin embargo, recientes encuestas han mostrado que las calificaciones en letras (A,B,C,D,F) se mantienen como la práctica más común de calificación en las escuelas primarias y secundarias” (Moll,1998: 1).

Es probable, como lo señala O'Connor (2010), que hayan sido las fuertes críticas surgidas a partir de los señeros estudios de confiabilidad realizados por Starch y Elliot en Estados Unidos y Laugier y Weinberg en Francia, las que llevaron a la adopción de escalas de calificación en letras masivamente ya hacia 1920.

En el caso de Argentina, la escuela secundaria estuvo marcada por el modelo del Nacional Buenos Aires fundado en 1863 con una fuerte impronta francesa. Una primera aproximación a las normativas de evaluación y calificación arroja que al menos desde 1947 la escuela media argentina mantiene casi sin cambios su escala de calificación de 1 a 10, el promedio (6 /7) como forma de acreditación de cada asignatura y la promoción

al curso superior con no más de dos asignaturas adeudadas. Las modificaciones establecidas durante el gobierno de Alfonsín tuvieron un fuerte rechazo en la comunidad educativa, lo que llevó a que –con un par de variantes- sólo estuviera vigente entre 1985 y 1990, aproximadamente. Su escala establecía 3 grados de aprobación (S-Superó; AMS-Alcanzó muy satisfactoriamente; A-Alcanzó) y un grado de desaprobación (ANS- Aún no satisfactorio). Esta escala procuraba una enseñanza y evaluación por objetivos que, volviendo a las calificaciones tradicionales, pervivió en el nivel medio ya descentralizado a partir de 1995.

3. El campo de la evaluación y el estudio de los sistemas de calificación

Decíamos más arriba que desde el mismo “comienzo” de las prácticas de calificación, se dieron críticas, ensayos, cambios e intentos de la más variada índole. Podemos hablar de otro momento bisagra, cuando entre 1913 y 1914 se publican las investigaciones sobre confiabilidad de Starch y Elliot en Estados Unidos y Laugier, Weinberg y Pieron a partir de la década del 20 en Francia. Starch y Elliot llevan adelante una experiencia de corrección múltiple en Inglés y Geometría, encontrando una enorme dispersión en las calificaciones asignadas por los profesores (Cureton, 1971; Pollio, 2004; O'Connor, 2010).

En el mundo anglófono la reacción frente a estas evidencias llevó a intentar una respuesta desde la medición educativa bajo la influencia de la psicometría y desde el supuesto que toda población refleja la distribución de una determinada capacidad representada en una curva normal. Así algunos defensores de este método incluso especificaban los porcentajes exactos de estudiantes a los que se habría de asignar cada calificación.

Por ejemplo el “modelo”: 6-22-44-22-6: siendo el primer 6 % el que recibe una F; luego 22 % D; el siguiente 22% C; otro 22 % B y finalmente el 6 % superior una A. Fundado primeramente en 1908 por Meyer, fue muy utilizado y aún hoy permea muchísimas prácticas de calificación en todo el mundo, ya sea en forma conciente y explícita o inconsciente e implícita (sin conocer su origen o reflexionar sobre su fuerte inconsistencia).

De hecho, el fenómeno que André Antibí llama “la constante macabra”: que un profesor debe tener un mínimo porcentaje de alumnos desaprobados sino quiere ser considerado laxo o directamente incompetente. (2012), tiene como supuesto también una

distribución teórica a priori que opera en el imaginario de los profesores y estudiantes de Francia y –me atrevería a decir, también en Argentina.

Los hallazgos de Starch y Elliot (1913) también llevaron a realizar otras investigaciones sobre los exámenes de Baccalauréat en Francia coordinadas por Laugier y Weinberg (1938). Luego, confirmados en diversas ocasiones por Piéron (1963) entre otros, concluyeron en “la insuficiencia radical de los métodos de evaluación utilizados corrientemente, ya que las notas asignadas a un mismo trabajo podrían presentar, según los correctores, una diferencia de 13 puntos sobre una escala de 20” (Cardinet, 1989: 42). Estas investigaciones abrieron lo que se dio en llamar luego la docimología crítica: un “análisis crítico del funcionamiento de la calificación y de los exámenes tradicionales” (Barbier, 1983: 49).

No se trata de cuestionar la existencia misma de las calificaciones o de los exámenes, pero apuntan a transparentar a realidad de su funcionamiento y de los objetivos que se proponen. También hay investigaciones que han desarrollado la comparación de las notas desde una mirada de género. Es uno de los hallazgos interesantes de Pierre Merle, otro destacado docimólogo (2007), que investiga los distintos factores que influyen en el sesgo de las calificaciones.

Todos los intentos que describe De Ketele (1993) en la profusa descripción de los distintos paradigmas de evaluación muestran la dedicación a la evaluación y el desarrollo de las distintas corrientes de evaluación formativa en las últimas décadas. Sin embargo, “la evaluación formativa no dispensa a los docentes de poner notas o redactar apreciaciones para informar a los padres, a la administración” (Perrenoud, P, 2008: 17) que permitan tomar decisiones de selección, orientación, promoción, acceso a determinadas universidades o becas. (Withney, D., 2006).

Díaz Barriga muestra a las calificaciones como un problema que el capitalismo ha introducido en la escuela, ya que “la asignación de una calificación es un problema de la institución educativa y de la sociedad, no una cuestión intrínseca a la pedagogía” (1994:176). Si hay sistema educativo, se necesita informar, comunicar, comparar, acreditar, responder (*accountability*) ante la sociedad y los otros niveles del sistema (universidad, por ejemplo).

No es la supresión de la nota, como señala el propio Maulini (1996) sin imaginar nada en su lugar, lo que va a ayudar a una evaluación más confiable y válida, que

deconstruya las inequidades y clasificaciones que han nacido con la propia escuela moderna. “No reivindicamos la supresión de la calificación. Es nuestra cultura de la evaluación y nuestra mentalidad la que debe ser cuestionada” (Antibi, A.,2012: 23).

Es un campo complejo, “entrecruzado” (Carlino,1999: 72) que debe ser abordado de manera de dar cuenta de todas sus perspectivas: pedagógicas, sociológicas, didácticas. Ya algunos especialistas han comenzado a pensar estrategias que permitan mejorar las prácticas de evaluación y las calificaciones en una perspectiva.

CAPÍTULO 3

LAS PRÁCTICAS DE CALIFICACIÓN DE LOS PROFESORES

Frente a la necesidad de revisar nuestras prácticas de calificación para ajustarlas a los marcos didácticos de una evaluación formativa conforme a los cambios que están implementando las políticas educativas según los lineamientos ya establecidos en la Ley de Educación Nacional 26.206 y en las resoluciones del Consejo Federal de Cultura y Educación 84/09 y 93/09 -que hemos citado en la introducción- consideramos fundamental indagar cuáles son las prácticas vigentes, las decisiones que toman los profesores del nivel secundario al momento de calificar, particularmente en las instancias sumativas del cierre de trimestre que se informan en el boletín de calificaciones. Estas instancias, en la normativa actual, van a ser determinantes para la nota final de acreditación de cada asignatura, la promoción al curso siguiente y/o la certificación final correspondiente al nivel medio en la orientación escogida.

¿Qué evalúa el profesor cuando califica? ¿Cuáles son los criterios que subyacen a sus prácticas de calificación? ¿Cómo valida sus decisiones? ¿Qué finalidades supone el uso que hace del sistema? ¿acreditar? ¿promover? ¿clasificar? ¿disciplinar? ¿Qué información ofrecen las notas producidas? ¿Qué tiene en cuenta el docente al momento de acreditar la asignatura? ¿Cómo intervienen los factores académicos y no-académicos (esfuerzo, dedicación, proceso de mejora) en la nota asignada?

Para intentar responder a estas preguntas, nos hemos planteado los siguientes objetivos de investigación:

- Objetivo general: Describir las prácticas de calificación que utilizan los profesores de nivel secundario -de las jurisdicciones de Ciudad de Buenos Aires y Provincia de Buenos Aires (gestión pública y privada) en las disciplinas de Historia, Lengua y Literatura y Ciencias Biológicas al cierre del trimestre.

- Objetivos específicos:

- a) Indagar qué criterios y operaciones ponen en marcha para decidir la nota del estudiante al cierre del trimestre.
- b) Identificar y analizar las dimensiones del desempeño del estudiante que toman en cuenta al decidir la nota de cierre.
- c) Determinar qué dimensiones de la validez subyacen a las prácticas de calificación de cierre del trimestre de los profesores de Lengua y

Literatura, Historia y Ciencias Biológicas.

Marco metodológico: test de escenarios de calificación de cierre de trimestre

La recolección de evidencia se llevó adelante sometiendo a una prueba de escenarios de calificación (adaptado y traducido a partir de la versión de Brookhart, 1993) de cierre del 2do. trimestre a 13 profesores de historia, 8 profesores de Biología y 7 profesores de Lengua y Literatura (gestión pública y privada) que se desempeñan en el nivel secundario de la Ciudad de Buenos Aires y/o en la jurisdicción de la Provincia de Buenos Aires.

Estos escenarios fueron utilizados originalmente para investigar qué dimensiones del desempeño de los alumnos utilizaban los profesores en las calificaciones (Manke y Loyd, 1990). Luego Susan Brookhart (1993) le incorporó una pregunta semiabierta - “¿Por qué razones elegiría esa alternativa?” - para indagar los criterios que subyacen en las opciones que eligieron los docentes en cada escenario.

Estos aspectos de las prácticas de calificación al final de un período trimestral responden adecuadamente a los objetivos arriba explicitados. Por otra parte, este mismo instrumento ha sido utilizado en China con profesores de inglés como idioma extranjero para indagar el significado que asocian con las notas que ponen a sus estudiantes y el juicio de valor que realizan al calificar (Sun, Y. y Cheng, L, 2014). Esto nos ha mostrado la vigencia del instrumento y nos permite asimismo comparar resultados en contextos culturales muy diferentes: Estados Unidos, China y Argentina.

Hemos incorporado 2 escenarios más a la versión original de 7, procurando introducir algunas situaciones que atraviesan las decisiones que toma el profesor en el cierre de la nota trimestral, que no figuraban en la versión original, a saber: un escenario de entrega tardía de trabajos prácticos y otro de desaprobación de la “prueba trimestral integradora” pese a haber dado cuenta del cumplimiento de los objetivos del período en una exposición oral.

Optamos por hacer referencia a situaciones de cierre del 2do trimestre ya que ha transcurrido el suficiente tiempo de conocimiento del grupo clase y no están aún fuertemente influenciadas por el proceso de acreditación de las asignaturas que se

realiza en el cierre del 3er trimestre y no pretendemos abordar.

La muestra es intencional, procurando una selección de casos que permita cubrir una importante heterogeneidad en edad (entre 18 y 59 años), sexo (15 mujeres y 13 varones), antigüedad en la docencia (entre 2 y 34 años) y formación de base (con título terciario docente, no-docente, universitario de grado y posgrado). En un enfoque microsociedad que mira a los actores protagónicos de las prácticas de calificación, tomaremos a este grupo de profesores como representativos del universo de docentes de nivel secundario de Historia, Lengua y Literatura y Ciencias Biológicas de sus jurisdicciones de pertenencia.

Para clasificar los comentarios realizados por los profesores a la pregunta semiabierta acerca del porqué de la opción elegida, hemos utilizado siguiendo a Brookhart (1993), el esquema de clasificación y puntuación basado en el modelo de validez de Messick -presentado en el capítulo 2 dedicado al marco teórico- que, identificando dos facetas de validez: la función intencional del puntaje (interpretación o uso) y la fuente de justificación (evidencia empírica o consecuencias sociales), forma una matriz progresiva cruzando estas dos facetas para formar una tabla del siguiente modo:

Figura 5: Matriz progresiva de validez.

Fuente de justificación	Función intencional	
	Interpretación	Uso
Evidencias	VC (validez de construcción)	VC + RU (relevancia y utilidad)
Consecuencias	VC + IV (implicancias de valor -axiológicas)	VC + RU + IV + CS (consecuencias sociales)

Fuente: Adaptado y traducido de Messick, 1989b.

Desde la perspectiva de Messick (1989a y 1989b) es la validez de construcción la que se complejiza según su interpretación y/o la utilización que se hace de los resultados de un test, o en nuestro caso de una nota (trimestral) sea que en las decisiones que toma el profesor pesen más o menos las evidencias empíricas que la sostienen y/o las consecuencias que pueda tener dicha calificación asignada a ese alumno en ese

contexto temporal (ese trimestre), espacial (esa escuela) y socio-cultural (su cultura, su barrio, su comunidad étnica, el sistema de estudios superiores, etc.). En definitiva “evaluar y calificar siempre se realiza en un contexto, no en el vacío” (Moss, C., 2013: 252).

La validez de construcción (VC) hace referencia al significado que el docente atribuye a la nota en sí misma: el contenido de la calificación, qué producciones del período tendrán una finalidad sumativa y, por lo tanto, serán calificadas y contabilizadas para la nota trimestral, así como su eventual ponderación. En esta dimensión tiene mucha fuerza el promedio invocado como criterio obligatorio por la normativa vigente y como eventual objetividad o neutralidad aparente que evita discusiones o cuestionamientos de actores externos al aula (directivos, profesores tutores, gabinete, padres). Otros aspectos que aparecen como parte de la nota son los facilitadores académicos (McMillan, 2001; 2019; Bonner, 2019) o factores de éxito escolar (Bowers, 2009) que no refieren a un desempeño o resultado de una determinada producción del estudiante. Entre éstos los autores suelen referir la asistencia, la actitud y atención en la clase, la puntualidad y el esfuerzo -entre otros-. Precisamente el esfuerzo parece ser muy valorado y parte importante de la nota trimestral según lo han expresado la mayoría de los profesores de esta muestra, aun cuando expliciten la dificultad de evaluarlo y/o calificarlo.

La relevancia y utilidad (RU) refiere al conjunto de evidencias específicas (Messick, 1989b) reunidas respecto del desempeño del estudiante concreto: sus resultados obtenidos, el esfuerzo realizado, notas previas u otras fuentes de evidencia que les permiten, sostenidos en este criterio, dar cuenta de la calificación informada en el boletín. Responde a la pregunta por el significado de esa nota cuando es asignada a un estudiante. Los profesores -en nuestro estudio- le dan mucha importancia a la evidencia ya que muchas veces quieren más información (incluyen más calificaciones o califican evaluaciones que, quizás, podrían ser más formativas que sumativas). Otras veces procuran extender el período de evaluación para poder incluir más evidencia que no han conseguido reunir o porque alguna instancia sumativa de mayor trascendencia en su planificación, en la institución (integradoras institucionales) o en el sistema (integradoras oficiales que pesan hasta el 50 % de la nota del tercer trimestre) ha arrojado un resultado muy disruptivo respecto de todo el proceso realizado por el alumno a lo largo del trimestre o del ciclo escolar. En general, de los comentarios que

hacen en los escenarios, surge que prefieren tomar sus decisiones con base en la evidencia (al menos la que han podido reunir o poseen de otras fuentes: otros profesores, departamento de orientación, otras asignaturas, etc.) y no por sus eventuales consecuencias, haciendo así hincapié especialmente en esa dimensión -en principio empírica- de las prácticas de calificación que llevan adelante. Incluimos bajo esta categoría aquellas evidencias del desempeño del docente que refieren en las explicaciones: los recordatorios que le han hecho al estudiante, cómo lo han alentado en el proceso o qué dispositivos le han ofrecido para poner en evidencia el aprendizaje. En este marco, dos profesores, ante el fracaso del alumno en una prueba integradora (escenario 7) señalan como posibilidad “rever la prueba que se tomó” o “revisar las herramientas que estoy utilizando para evaluar” (escenario 6). Si bien suponen consecuencias ulteriores, no son proyectadas en el estudiante, sino en la reflexión del docente sobre su propia práctica. Sino hubiésemos clasificado estos comentarios dentro de CS. Por cierto, fueron sólo dos docentes en dos escenarios sobre 252 comentarios entre 28 docentes.

Cuando se abordan las consecuencias que tiene la interpretación de las notas asignadas aparecen implicancias de valor o axiológicas (IV). Esta dimensión responde a la pregunta por los valores que subyacen a las decisiones tomadas por el docente al calificar a un estudiante. Como en las investigaciones de Brookhart (1993) y de Sun y Cheng (2014) la palabra más utilizada en este tipo de comentarios fue “justo” o “injusto”. Muchas expresiones, en particular en los escenarios 9 y 1 dan cuenta de la preocupación por tratar a todos los alumnos equitativamente, reforzando la idea de hacerlo en forma consistente con todos según las pautas de evaluación y calificación previamente informadas y acordadas con la clase. Otros que, al igual que Brookhart (1993) también clasificamos entre los que ponen en juego aspectos axiológicos, tienen en cuenta dificultades personales y/o de aprendizaje. Algunos docentes, incluso, pese a haber tomado una determinada decisión de calificación, se preguntan si, por ejemplo, una caída en el rendimiento de un alumno no habría obedecido a una dificultad de aprendizaje o situación familiar delicada. Son consideraciones que podemos llamar de compasión (Brookhart, 1993) o empáticas.

En la matriz de Messick (1989b) también se incluye la consideración de las consecuencias sociales del uso de las calificaciones (CS). Esta dimensión, además del significado de la nota asignada a un estudiante y los valores que subyacen a esa decisión,

considera qué sucederá a causa de esa nota. Las consecuencias que puede tener a futuro el uso de esa calificación informada tanto en sus trayectorias escolares (extensión del tiempo de enseñanza, aprendizaje y evaluación, acreditación, promoción, certificación), sus itinerarios académicos en estudios superiores y/o en la vida personal o profesional del estudiante. De algún modo el docente ha tenido en cuenta las consecuencias “a futuro” que derivan de la nota asignada a ese estudiante tanto en el ámbito escolar como extraescolar. Prevé cómo impacta la nota en la perseverancia y mejora continua que afectará su desempeño ulterior en la escuela en las diversas asignaturas y, también, otras consecuencias extraescolares que deriven en cuestiones inmediatas de repercusión de las decisiones del docente ante los padres o en otros aspectos más a largo plazo que generen aprendizajes relevantes del estudiante para su vida personal o profesional: aprender a rendir cuentas de su trabajo, fortalecer su autoestima o evitar frustraciones que lo bloqueen para el estudio de la asignatura evaluada u opciones vocacionales ligadas a ella. Algunos comentarios que realizaron los docentes de la muestra refieren una suerte de acuerdo o pacto -cuyo contenido sólo queda en la conversación privada estudiante-profesor o en algunos casos ante toda la clase- de “préstamo de un punto” que también se proyecta a futuro porque derivará en sus calificaciones de la asignatura en los siguientes trimestres. Este tipo de arreglos -como los llama y describe ampliamente Pierre Merle (2007)- son frecuentes en las prácticas de calificación, aunque hayan surgido en las explicaciones de los docentes sólo éstos referidos al “préstamo de un punto”. No cabe duda de que hay por un lado un trabajo que podríamos llamar individualizado con el alumno que, a la vez conlleva un deterioro de la validez de construcción en tanto parecería sugerir una interpretación totalmente específica y coyuntural de esa nota asignada a ese alumno en ese momento particular. Sin embargo, esta situación nos abre a la consideración contextual, idiosincrática y cultural que no vamos a profundizar en este análisis, pero de la que dan cuenta investigadores de las más diversas corrientes: desde Sverdlick (2012) a Moss (2013), pasando por Bonner que nos recuerda que “las notas cuantifican atributos o construcciones teóricas cuyos significados son negociados social o culturalmente” (2019: 57).

Siguiendo el modo de puntuar de Brookhart (1993) y dado que la matriz de Messick (1989b) es progresiva e indica un enriquecimiento cualitativo de los fundamentos de la validez que incluye las dimensiones anteriores, otorgamos a los comentarios ligados a la validez de construcción (VC) 1 punto, a los de relevancia y

utilidad (RU) 2, a los que implican valores (IV) 3 y a los que consideran consecuencias sociales (CS) 4, respectivamente.

Organización y análisis de los datos cuantitativos y sus criterios de validación

Vamos a analizar en primer lugar las decisiones en la expresión estereotipada que propone el escenario para luego establecer la validación de dichas prácticas de calificación según la matriz de validez dado que diversas investigaciones dan cuenta de la necesidad de dar cuenta de los motivos que fundan las decisiones de calificación que toman los profesores (Brookhart, 1993; Guskey, 2018). Tal como lo hemos explicado, clasificaremos las justificaciones de cada elección realizada por los profesores en cada escenario del siguiente modo, siguiendo la propuesta original de Susan Brookhart (1993), a saber:

- 1) Validez de construcción (VC) ¿Qué significa esta calificación en sí misma?
- 2) Relevancia y utilidad (RU) ¿Qué significa esta calificación cuando es asignada a un alumno?
- 3) Implicancias de Valor (IV) ¿Qué significa la nota cuando es asignada a un estudiante y qué valor tiene?
- 4) Consecuencias sociales (CS) ¿Qué significa la nota cuando es asignada a un estudiante, cuál es su valor y qué sucederá a causa de ella?

Para clasificar los distintos comentarios realicé el chequeo con otro investigador y los desacuerdos fueron resueltos por consenso.

Figura 6. Ejemplos de las categorías de la matriz de validez y sus comentarios

Categoría	Ejemplos de comentarios	Explicación
VC	Dejo en claro la forma de evaluación al principio de cada año y trimestre y, para ser coherente con lo dicho, respeto las mismas.	Refieren los criterios de evaluación y calificación que han sido explicitados y/o que indica la norma vigente.
	Si la tarea es parte de la nota, tiene que contar.	Indican cuáles son las producciones que reciben calificación.
	No sólo tomo pruebas.	
	También incluyo una nota por esfuerzo, participación en clase, carpeta.	Explican qué significa la nota: pago por el trabajo

Si se había pautado que el trabajo práctico tenía un peso del 50 % hay que cumplirlo. hecho, un puntaje calculado, un promedio.

La nota del trimestre refleja el trabajo y esfuerzo realizado.

De ninguna manera calificaría algo que no hizo.

Es lo que le dio el promedio.

Si se sacó un 6, entonces esa es su nota.

Si el 9 es merecido se lo pongo.

Estuvo mal ponderado el promedio de la evaluación integradora

RU	<p>Quiero saber su nota del 1er trimestre. Reveo la prueba que se tomó.</p> <p>Tiene capacidad para reconocer que su esfuerzo no ha sido el de siempre.</p> <p>Debería tomar un recuperatorio antes de cerrar el trimestre.</p> <p>Además, se le recordó insistentemente.</p> <p>Quisiera tomarle un recuperatorio y/o incluir nota de su carpeta... como suelo hacer con mis alumnos.</p>	<p>Expresan las evidencias del desempeño de este estudiante en particular y las calificaciones obtenidas y oportunamente comunicadas al estudiante y/o a los padres.</p> <p>Incluso en otras instancias o asignaturas o requieren indagar sobre su situación personal a través de directivos o departamento de orientación.</p> <p>Expresa las evidencias del desempeño del docente y cómo ha sido su trabajo con ese estudiante.</p>
IV	<p>Estoy evaluando responsabilidad y no sería justo con el resto del alumnado.</p> <p>No sería justa con los alumnos que si cumplieron.</p>	<p>Los comentarios apelan a valores que subyacen a las decisiones de calificación. En particular hacen</p>

	<p>Es importante el detalle de que fuera advertido y avisado</p> <p>Califico también lo actitudinal, aunque el resultado académico es bueno. Su calificación, por lo tanto, es más baja.</p> <p>Me parece justo.</p> <p>Se le brindaron varias oportunidades.</p> <p>Siempre un refuerzo positivo nos da mejores resultados</p> <p>Tengo en cuenta si está pasando por una situación personal o familiar.</p>	<p>referencia a la equidad (aunque lo llaman justo en la mayoría de los casos), a las oportunidades, a lo correcto o incorrecto.</p> <p>Apelan también a la compasión, empatía o misericordia ante problemas de los estudiantes.</p> <p>Apelan a creencias del docente.</p>
CS	<p>Si le pongo un 5 sólo logro que se frustre o bloquee o abandone la materia.</p> <p>Es fundamental incentivar y promover el esfuerzo personal del alumno.</p> <p>No sólo porque así se pautó sino porque le sirve para la vida.</p> <p>Cumplir en tiempo y forma lo prepara para la facultad.</p> <p>No elijo la opción de bajarle la nota porque sería muy difícil la justificación ante los padres si eventualmente decidieran quejarse</p> <p>En el tercer trimestre habrá oportunidad de “ayudarla” si es que la situación lo requiere y hablaría con ella para alentarla a que siga esforzándose. Pero esta nota le haría notar a ella y a su familia que quizás necesita otra ayuda externa.</p> <p>Considero que podré ver avances en él durante el tercer trimestre, seguramente</p>	<p>Consideran las consecuencias, mira hacia el futuro del estudiante.</p> <p>Consecuencias escolares: el desempeño o actitud del estudiante al continuar el cursado de la asignatura, en el siguiente trimestre o año escolar o en otras asignaturas. Procuran evitar quejas de los padres o los involucran en el desempeño de los estudiantes.</p> <p>Consecuencias extraescolares ligadas a habilidades para la vida relacionadas con la actitud, el esfuerzo, el sostener la autoestima y aprender a rendir cuentas.</p>

Fuente: Elaboración propia (adaptando formulaciones de Brookhart, 1993 y de Sun y Cheng, 2014)

Escenarios de calificación: respuestas, comentarios y análisis.

Cuadro 1: Escenario 1 de trabajo de acuerdo con la capacidad

Escenario 1

Usted es profesor de un curso de 1er año. Tiene que realizar el cierre del segundo trimestre. Cristina, una de sus alumnas tiene una gran capacidad académica, como lo ha demostrado en sus trabajos anteriores, resultado de las pruebas, informes de otros profesores y su propia observación.

Cuando revisa sus trabajos para el período, se da cuenta de dos cosas:

La calidad de sus trabajos es superior al promedio del curso, pero no reflejan el nivel que Cristina puede lograr. Usted considera que el esfuerzo que ha hecho es mínimo, pero, por su gran capacidad, el resultado es razonablemente bueno.

En esta situación, usted:

N/28	%	Opciones ofrecidas por el escenario
20	71.4	a) Califica a Cristina por la calidad de su trabajo comparado con el resto de la clase sin preocuparse por el nivel que podría haber alcanzado
6	21.4	b) Le baja la nota a Cristina porque no hizo un esfuerzo serio. Podría haber hecho un trabajo mejor.
2	7.14	c) Le sube la nota a Cristina para alentarla a trabajar mejor en el próximo trimestre.

¿Por qué elegiría esta opción?

Opción	Puntaje de validez progresiva				Total
	VC = 1	RU = 2	IV = 3	CS = 4	
Por calidad	10	4	4	2	20

Bajar la nota	2	2	2		6
Subir la nota	1			1	2
Total	13	6	6	3	28

Moda: VC (1)- 46.4 %

Fuente: Adaptado de Brookhart, 1993.

En este primero y en el segundo escenario, analizando las decisiones que prefiere tomar cada profesor la mayoría elige (75 % en el nro. 1 y 92.8 % en el nro.2) la opción de calificar de acuerdo con la calidad del trabajo realizado independientemente del desempeño previo o de su conocida capacidad. Se evidencia una gran proporción de docentes que prefiere atenerse a una nota que tenga directa relación con las calificaciones parciales que han comunicado previamente.

Brookhart (1993), para los docentes en ejercicio que están haciendo estudios de posgrado en la Universidad de Duquesne (Pennsylvania – Estados Unidos) encuentra resultados similares: fueron un 68.8 % aquellos que optaron por asignarle a la estudiante la nota de acuerdo con su desempeño, más allá de su potencial. Por otra parte, en este mismo escenario, los profesores de idioma inglés en China (Sun y Cheng, 2014) optan en un 42,1 % por bajarle la nota que hubiera obtenido por sus desempeños al estudiante que ha rendido por debajo de su capacidad.

A la hora de analizar los motivos que esgrime cada docente para realizar dicha opción, podemos observar que la fundamentación de la validez de sus elecciones, de acuerdo con la matriz de validación de Messick. Asignando un puntaje progresivo a cada una de las razones que fundan la decisión tomada, podemos observar lo siguiente:

En los comentarios vemos que la fundamentación más frecuente estuvo sostenida en la validez de construcción (VC): “la nota es acorde a su trabajo” (Test 16), “califico según los objetivos planteados en el programa” (Test 14). Incluso los docentes que han optado por bajar la nota de Cristina señalan que la nota “no debe tener un carácter punitivo”, aunque es importante – dice el mismo docente- “motivar el talento” (Test 18).

Entre los que fundamentaron su decisión en criterios de relevancia y utilidad (RU) señalan algunos la necesidad de conocimiento acerca de algún problema personal

que haya llevado a la alumna a rendir por debajo de su capacidad (Test 20) y sostienen su práctica de calificación en la evidencia disponible ya que, “de ninguna manera calificaría por algo que no hizo” (Test 27).

Una docente consideró bajar la nota a Cristina fundada en cuestiones axiológicas que considera importante resaltar: “Califico también lo actitudinal, aunque lo académico sea bueno” (Test 24), en tanto que otro señala que le “parece justo – dejarle la nota obtenida en sus desempeños- aunque no haya trabajado conforme a su potencial” (Test 11).

Desde una mirada dirigida a las consecuencias que tiene el uso e interpretación de la nota, una profesora hubiese subido la nota a Cristina “haciendo explícito que no debe desaprovechar su potencial, sobre todo teniendo en cuenta que está en primer año” (Test 5).

Un aspecto que vamos a encontrar en varios escenarios es la necesidad que expresan los docentes de dialogar con el estudiante (Test 10) e intentar otras estrategias, aunque la decisión de calificación haya reflejado un ajuste estricto a sus desempeños. Pareciera que hay aspectos pedagógicos que el docente quiere tener en cuenta, pero considera que no deben integrar la calificación o ésta sólo permite en su síntesis sumativa una única información y prefiere atenerse a la normativa.

Cuadro 2: Escenario 2 de trabajo de acuerdo con la capacidad.

Escenario 2

Usted es profesor de 2do año. A Bárbara le cuesta aprender, según se observa por su desempeño actual y la información que le transmiten sus anteriores profesores.

Usted se da cuenta de que ella ha trabajado con mucho esfuerzo durante el trimestre. Entregó todos los trabajos entiendo y forma y se ha acercado frecuentemente a pedir ayuda extra antes de las pruebas.

Su promedio del segundo trimestre –según las notas que tiene- es un $5/6^2$. En esta

² En la Ciudad de Buenos Aires 6 es la nota de aprobación para el promedio general anual de cada asignatura entre los tres trimestres y al menos 6 en el tercero. En la Provincia de Buenos Aires es 7 la nota

situación Ud.:

N/28	%	Opciones ofrecidas por el escenario			
26	92.8	a)	Le sube la nota a un 6/7 por el esfuerzo realizado.		
2	7.14	b)	Le deja el promedio 5/6		
		¿Por qué elegiría esa opción?			
Escenario 2		Puntaje de validez progresiva			
Opción	VC = 1	RU = 2	IV = 3	CS = 4	Total
Sube la nota	2	8	9	7	26
Deja la nota			1	1	2
Total	2	8	10	8	28

Moda: 3 (IV)

Fuente: Adaptado de Brookhart, 1993.

En este segundo escenario de trabajo de acuerdo con la capacidad, los docentes se inclinaron en notable mayoría por valorar el esfuerzo de la estudiante en función del conocimiento que tienen de sus dificultades de aprendizaje (aunque el escenario no supone una específica adaptación curricular fruto de un diagnóstico explícito que hubiere derivado en un proyecto de inclusión bajo la resolución 1644/17 de la Provincia de Buenos Aires, por ejemplo). Aquí vemos el fenómeno que McMillan (2019) describe cómo “apoyar a los estudiantes” - *pulling for students* -. Es uno de los aspectos que genera cierta confusión en la información que brinda la nota ya que no queda claro cómo y por qué a un determinado estudiante se le considera el factor esfuerzo o mejora. Por ello dice que es un tema que hay que discutir entre los docentes, ya que “sería de ayuda

de aprobación para el promedio general de cada asignatura y al menos 4 en el tercero. Por eso según el docente se desempeña en una u otra jurisdicción en este escenario se adaptó la nota de aprobación

proveerle a esta expresión: “apoyo a los estudiantes” de una clara fundamentación racional para operacionalizarlo y, así, asegurar mejor la justicia y consistencia en y entre los grupos de clase” (McMillan, J., 2019: 105).

Los valores que entran en juego en la decisión que tomaron muchos docentes (IV) apelan a cierta actitud empática de solidarizarse con la dificultad y el esfuerzo que realizó la estudiante a pesar de no lograr el desempeño académico esperado. “Siempre un refuerzo positivo genera mejores resultados” (Test 4). De entre los poquísimos docentes que prefirieron dejar el promedio “que correspondía”, llama la atención que su fundamentación haya sido también ligada a los valores (IV): “siempre brindo las oportunidades necesarias para que aprueben con base en sus capacidades” (Test 28) y a las consecuencias a futuro (CS): “En el tercer trimestre habrá oportunidad de “ayudarla” si es que la situación lo requiere y hablaría con ella para alentarla a que siga esforzándose. Pero esta nota le haría notar a ella y a su familia que quizás necesita otra ayuda externa” (Test 25).

Cuadro 3: Escenario 3 de trabajo de acuerdo con la capacidad.

Escenario 3

Usted es profesor de un curso de 1er año. Sandra, una de sus alumnas, es una alumna promedio. Al analizar sus trabajos de este trimestre, observa que podría haber trabajado mejor, dada su capacidad. Sin embargo, el promedio le da 6.

En esta situación, Ud.:

N/28	%	Opciones ofrecidas por el escenario
3	10.7	a) Califica a Sandra con una nota más baja porque se podría haber esforzado más y hacer mejores trabajos.
24	85.7	b) Califica a Sandra de acuerdo con la calidad de sus trabajos sin tener en cuenta que podría haberse esforzado más.
1	3.5	c) Califica a Sandra con una nota más alta para alentarla a esforzarse más.

		¿Por qué elegiría esta opción?			
Escenario 3		Puntaje de validez progresiva			
Opción	VC = 1	RU = 2	IV = 3	CS = 4	Total
Baja la nota	1	1	-	-	2*
Por calidad	12	2	4	5	23*
Sube la nota	-	-	1		1
Total	13	3	5	5	26

Moda: VC (1) – 50 %

Fuente: Adaptado de Brookhart, 1993.

Nota: Un docente que hizo la opción b y otro que hizo la opción a no justificaron su decisión.

En este último escenario que plantea la situación de un estudiante que se ha desempeñado por debajo de sus posibilidades –“podría haberse esforzado más”- una importante mayoría de docentes optó por calificarlo de acuerdo con su desempeño. Del mismo modo surge de la muestra analizada en Pennsylvania por Brookhart (1993) con una aún más marcada tendencia (95 %) a calificar ajustándose a la calidad del trabajo que ha realizado la alumna. El estudio de Sun y Cheng (2014) no incluyó este escenario.

El 50 % de los docentes anclan sus decisiones en la validez de construcción (VC), prefiriendo atenerse a los resultados obtenidos, aunque notan el bajo esfuerzo que ha puesto la estudiante en el trimestre. Sin embargo, casi todos insisten en hablar con la estudiante y/o reconsiderar el bajarle la calificación en el trimestre posterior si continúa mostrando una pobre dedicación (Test 23). Nuevamente parece ser insuficiente la calificación para la información que quiere transmitir el profesor en la evaluación sumativa trimestral que necesita complementar de otro modo.

De entre los que argumentan desde las consecuencias (CS), una docente prefiere – en el escenario 1 y en el 3- dejar la nota ajustada a la calidad del desempeño de la

estudiante porque prefiere evitar posibles conflictos con los padres si le baja la calificación (Test 3). Quizás los docentes, aunque desde una mirada de evaluación formativa alientan el esfuerzo, la dedicación y que cada estudiante dé lo mejor de sí de acuerdo con sus capacidades, parecen verse encerrados en el “callejón sin salida” de la evaluación sumativa que sintetiza sólo resultados que estén ligados a objetivos definidos (CV) y/o evidencia del trabajo realizado por el estudiante.

Cuadro 4: Escenario 4 de calificación en caso de trabajos no entregados.

Escenario 4

En su clase de 2do año, para la nota del segundo trimestre Ud. tomó en cuenta las diferentes evaluaciones escritas y un trabajo práctico de investigación para hacer en casa -cuya calificación sería el 50 % de la nota del trimestre-.

Tomás se sacó un 9 de promedio en las 3 evaluaciones escritas, pero no entregó el trabajo práctico a pesar de recordárselo varias veces. En esta situación, Ud.:

N/27	%	Opciones ofrecidas por el escenario
-----	-----	a) No considera el TP sin presentar y le pone un 9 a Tomás en el segundo trimestre.
25	92.5	b) Le pone un 1 por el trabajo no entregado y le desaprueba el segundo trimestre porque su promedio hubiese sido 5.
2	7.4	c) Le baja un poco la nota (el 9) del segundo trimestre por no haber entregado el proyecto.

En las opciones elegidas, los profesores testeados en este estudio se inclinan siempre por respetar las pautas y criterios que habían establecido con su grupo de clase y se ajustan a él estrictamente. Sólo un 7,4 % considera una posibilidad levemente más flexible. Los docentes de Pennsylvania (Brookhart, 1993), en tanto, se distribuyeron casi por partes iguales entre las opciones b y c: 51 y 48 %, respectivamente. Es decir, casi todos- salvo un caso- incluyeron el trabajo no entregado (con la nota correspondiente) en

la evaluación sumativa trimestral, aunque hubo una distribución más homogénea en el modo de realizarlo que en las opciones de los profesores de nuestro testeo. Sun y Cheng (2014) no utilizaron este escenario.

Veamos como justificaron sus decisiones de calificación:

Escenario 4	Puntaje de validez progresiva				Total
	VC = 1	RU = 2	IV = 3	CS = 4	
Opción					
No considera	-----	-----			
Promedia TP no entregado	11	7	6	1	25
Baja un poco por TP no entregado	1	-----	-----	1	2
Total	12	7	6	2	27

Moda: 1 (VC) – 44 %

Fuente: Adaptado de Brookhart, 1993.

Apoyaron su elección en la validez de construcción que no sólo implica la perspectiva de una adecuada interpretabilidad de la nota en cuanto promedio, sino que ésta incluye, a su vez, la ponderación del trabajo práctico que estableció el docente como instancia obligatoria con un peso del 50 % como lo indica en el escenario. “Ellos sabían que esa nota tenía “el peso” del 50 % en la calificación final” (Test 7). En este sentido consideran parte de la validez de construcción el hecho de transparentar e informar claramente los criterios de evaluación y calificación. “Dejo en claro la forma de evaluación al principio de cada año y trimestre y, para ser coherente con lo dicho, respeto las mismas. Esto no quita que charle con el alumno y, si la situación lo amerita con los padres” (Test 29). Esto garantiza interpretabilidad de la nota sostenida prioritariamente desde la “línea horizontal” de las evidencias en la matriz de Messick (1989b). Como en otros escenarios.

“Encima se le recordó varias veces que debía entregarlo” (Test 8), lo que implica una concreta significación de la nota para ese alumno en concreto (RU) que le da aún más

fuerza a la transparencia de la calificación sostenida en evidencia respecto de la comunicación al grupo y al alumno -en particular- por parte del docente.

Cuadro 5: Escenario 5 de calificación en caso de trabajos no entregados.

Escenario 5

Ud. es un profesor de un curso de 2do año. Para este trimestre la nota final de los alumnos se basa en las diferentes evaluaciones escritas y ejercicios para hacer en casa. José no entregó ninguno de estos ejercicios a pesar de sus frecuentes reclamos. Sus notas en las evaluaciones cortas estuvieron entre 6.50 /7.50 y 7.50/8.50³ y se sacó un 6 en cada una de las pruebas integradoras.

N/28	%	Opciones ofrecidas por el escenario
25	89.3	a) Ud. le promedia un 1 por incumplimientos.
3	10.7	b) Ignora los incumplimientos de José y le pone un 6/7 de promedio.
0	0	c) Ignora los incumplimientos y califica a José con un promedio de 7/8. ¿Por qué elegiría esta opción?

Escenario 5	Puntaje de validez progresiva				
	VC = 1	RU = 2	IV = 3	CS = 4	Total
Promediar el 1 por incumplimientos	10	5	8	2	25
Ignorar y poner una nota de aprobación.	1	---	1	1	3
Ignorar y calificar	---	---	---	---	---

³ Tanto en el escenario como en las opciones hay dos números en cada dato porque los escenarios para docentes de la Ciudad de Buenos Aires tuvieron en cuenta que la nota de aprobación necesaria (nota de corte) es de 6 y que en el caso de la Provincia de Buenos Aires es de 7.

con una nota mayor.				
Total	11	5	9	3

Moda: VC (1) – 39.2 %

Fuente: Adaptado de Brookhart, 1993.

Teniendo en cuenta que en el otro escenario de trabajo no entregado (N°4) tuvimos 27 respuestas válidas y en éste 28, los resultados serían exactamente los mismos, aunque hubo algunas variantes en la situación descripta para José y para Tomás, como así también en las características de la decisión de calificación ofrecida en las opciones. Habíamos visto que en los escenarios que presenta Brookhart (1993), en el anterior escenario de trabajo no entregado, había una distribución homogénea entre las dos opciones que proponían, de modo diferente, incluir la falta de entrega en la nota trimestral. Sin embargo, en este escenario que tiene una similar situación han optado mayoritariamente -un 86%- por promediar dicho incumplimiento sin más, como lo hicieron los docentes argentinos de esta muestra en un 89.3 %. En el caso de los docentes que relevaron Sun y Cheng (2014) encontraron una distribución bastante homogénea entre las 3 opciones (43.7 % para a; 32.2 % para b y 24.1 % para c) con mayor incidencia de la primera, como en Estados Unidos y Argentina, pero mucho menor globalmente. Los comentarios hechos por los profesores variaron bastante ya que se inclinaron más a valorar la ejercitación de tarea para el hogar como relevante (RU) y ponderaron que su incumplimiento refleja la actitud del estudiante ante la asignatura y sus hábitos de aprendizaje. Me parece interesante señalar que la muestra de esta investigación es sólo de profesores de inglés, la de Estados Unidos de distintos docentes de los que no tenemos la asignatura explicitada y, en nuestro caso, hay docentes de 3 asignaturas: Historia, Lengua y Biología. No está reflejado en el trabajo de campo, pero mi experiencia me dice que -al menos en las jurisdicciones de referencia- los docentes de idioma extranjero suelen darle mucha trascendencia a la ejercitación en la casa, aunque luego sea problemática su integración en una evaluación sumativa frente a un desempeño aceptable o bueno en las instancias de comprobaciones breves o pruebas integradoras.

Dado que los comentarios que orientan la validez hacia la construcción y hacia la relevancia tienen prácticamente la misma distribución en ambos escenarios, vamos a traer a colación algunas respuestas de implicancia axiológica (IV) que han aumentado en este

escenario con un 32 %, respecto del anterior en el que dicha categoría de validez estaba representada por un 22 %. “Estoy evaluando responsabilidad y no sería justo con el resto del alumnado” (Test 11), “no sería justa con los alumnos que si cumplieron. Es importante el detalle de que fuera advertido y avisado” (Test 1). Ambas explicaciones apuntando a la equidad en el trato entre estudiantes y a la consistencia de una pauta informada y recordada.

El siguiente escenario también refleja un tipo particular de trabajo no entregado que es su entrega tardía *-late work-*, procurando, así, indagar la eventual importancia que tiene dicho aspecto respecto de otros antecedentes del estudiante, en las prácticas de calificación de los docentes de la muestra.

Cuadro 6: Escenario 6 de calificación en caso de trabajos no entregados.

Escenario 6

Ud. es profesor de 3er año en un colegio secundario de la Ciudad de Buenos Aires. Al cerrar el segundo trimestre se encuentra que Juan Pablo – un alumno de habitual muy buen desempeño y cumplimiento- obtendría un promedio de $5/6^4$ de acuerdo con las calificaciones de sus pruebas, lecciones orales y trabajos prácticos presentados. Algunos incluso han sido presentados completos, pero en forma tardía. Por otra parte, el promedio general del curso es casi de $7/8$. En este caso,

N/27	%	Opciones ofrecidas por el escenario
18	66.6	a) Ud. deja el promedio $5/6$ a Juan Pablo
1	3.7	b) Le baja el promedio a $4/5$, ya que todo el curso ha rendido muy bien y el no.
8	29.6	c) Le sube el promedio, colocándole una nota elevada de concepto para alentarlo a mejorar y recuperar su habitual desempeño.

⁴ Al igual que en el escenario anterior las dos opciones, por ej:

¿Por qué elegiría esta opción?

Escenario 6	Puntaje de validez progresiva				
	VC = 1	RU = 2	IV = 3	CS = 4	Total
Opción					
Deja el promedio	5	5	5	3	18
Baja el promedio	----	----	----	1	1
Sube el promedio (y lo aprueba así)	1	2	3	2	8
Total	6	7	8	6	27

Moda: 3 (IV) – 29.26 %

Fuente: Elaboración propia.

En este escenario, si bien una mayoría del 66.6 % se inclina por ajustarse al promedio que viene obteniendo el alumno, es mucho menos marcada la tendencia a respetar las pautas prescritas (en este caso la entrega en tiempo) que en el caso de los otros escenarios en los que el estudiante directamente no había cumplido con la entrega que, en el escenario 4 implicaba el 50 % de la nota trimestral – allí un abrumador 92.5 % optó simplemente por promediar el 1 por incumplimiento – o bien los ejercicios para realizar en su casa que estaba pautado incluir en la nota y que un 89.3 % de los profesores deciden promediar. Pareciera -desde lo relevado- que la entrega tardía no debiera bajar el promedio que el alumno obtiene en sus desempeños ni tampoco perjudicarlo la comparación con un rendimiento global del curso respectivo

Sin embargo, un grupo relativamente importante de un casi 30 % parece tener presente la historia académica del alumno “de habitual muy buen desempeño y cumplimiento” y la posibilidad de alentarlo.

Lo que es sumamente variado y rico, y nos muestra como la complejidad de los escenarios deriva en prácticas similares, pero con criterios diferentes, es la notable dispersión en la clasificación de los comentarios según su validez de uso o interpretación.

Entre los que optaron por dejarle el promedio algunos consideran que el hecho de desaprobado lo ha de ayudar en el futuro a mejorar su rendimiento (Test 14) con consecuencias en su itinerario escolar (CS). El docente explicita que no quiere dar un mensaje de que “esté haciendo las cosas de una manera correcta” (Test 14), preocupado por la interpretación que el propio estudiante hiciera de una nota de aprobación trimestral que no refleje sus desempeños parciales por debajo del nivel pedido.

Los que se inclinaron por una justificación en valores (IV) siguen preocupados por “ser justo frente a los demás estudiantes que se esforzaron al máximo para obtener notas superiores” (Test 29) o “no modificar las reales calificaciones de los estudiantes que no presentan trabajos en tiempo y forma, por respeto a los que sí los entregan” (Test 5).

Uno de los aspectos que suelen incluirse en las prácticas de calificación de la nota trimestral tiene que ver con el proceso de aprendizaje que ha realizado el estudiante que ha ido mejorando y que, muchas veces, si se promedia estrictamente, no suele reflejarse en la nota del boletín. Dos escenarios nos plantean esa situación.

Cuadro 7: Escenario 7 de situaciones de mejora.

Escenario 7

Ud. es profesor en un secundario de la Ciudad. En su clase toma dos evaluaciones escritas por cada trimestre. David se sacó un 4.50/ 5.50 en la primera. En la segunda se sacó un 6 /7. En esta situación:

N/26*	%	Opciones ofrecidas por el escenario
22	84.6	a) Califica a David con un 5/6 en el segundo trimestre basándose en el promedio de su desempeño en los dos exámenes.
4	15.3	b) Califica a David con un 6/7 en el segundo trimestre porque mejoró su desempeño.

¿Por qué elegiría esta opción?

Escenario 7

Puntaje de validez progresiva

Opción	VC = 1	RU = 2	IV = 3	CS = 4	Total
Basado en el promedio	11	6	1	3	21
Basado en la mejora	2	2	----	----	4
Total	13	8	1	3	25*

Moda: 1 (VC) 52 %

Fuente: Adaptado de Brookhart, 1993.

*Un docente no comentó su decisión.

Realmente son muy pocos los que optan por apoyar la mejora del estudiante, ya que se ajustan al promedio de su desempeño sin considerar otra posibilidad como brindarle una nota de concepto o de progreso o que incorpore producciones que realmente evidencien el progreso que se ha visto en la segunda prueba escrita.

Cuadro 8: Escenario 8 de situaciones de mejora.

Escenario 8

Ud. es profesor en un curso de una escuela secundaria de la Provincia de Buenos Aires. En este curso Ud. toma dos exámenes por trimestre. Al cerrar las notas del segundo trimestre de Bernardo, observa que en la primera prueba se sacó un 7/8⁵ y en la segunda prueba un 9/10. En esta situación, Ud.:

N/28	%	Opciones ofrecidas por el escenario
25	89.2	a) Califica a Bernardo con un 8/9, promedio del puntaje de sus dos exámenes.
3	10.7	b) Califica a Bernardo con un 9/10, para señalar la mejoría de su desempeño.

⁵ Del mismo modo que en otros escenarios la nota que figura en primer lugar es la que utilizamos en los escenarios aplicados a docentes de Ciudad y la que figura en segundo lugar la que se usó para los escenarios para docentes de Provincia de Buenos Aires. Siempre teniendo en cuenta que la nota de aprobación es 6 en Ciudad y 7 en Provincia.

¿Por qué elegiría esta opción?

Escenario 8	Puntaje de validez progresiva				Total
	VC = 1	RU = 2	IV = 3	CS = 4	
Basado en el promedio	13	4	6	-----	23
Basado en la mejora	2	1	-----	-----	3
Total	15	5	6	-----	26*

Moda: 1 (VC) – 57.6 %.

Fuente: Adaptado de Brookhart, 1993.

*Dos docentes no justificaron su opción.

Hemos optado por analizar los resultados de los dos escenarios de mejora en su conjunto, ya que la investigación de Brookhart (1993) informa -especialmente en el escenario 8- de una inclinación a apoyar la mejora, calificando con una nota que refleje la misma en la evaluación sumativa trimestral. Hay dos observaciones que me parecen muy interesantes, al respecto: a) Ella nota- teniendo en cuenta el marco en el que hace su investigación- que han sido más propensos a tener en cuenta la mejora los docentes que han recibido mayor capacitación en medición y evaluación -un aspecto que quería investigar- ; b) Asimismo verifica que en ambos casos los docentes que se apartaron del estricto promedio para apoyar la mejora del estudiante se inclinaron más que en los escenarios anteriores (de calificación según capacidad y de trabajos no entregados) a justificar su decisión en argumentos de validez que ponderan las consecuencias ulteriores de dicha opción para el alumno.

Precisamente vemos una tendencia opuesta, más allá de las diferencias de los estudios, en los docentes que participaron de esta muestra. No solamente fueron muy consistentes en ajustarse al promedio, sino que además muy pocos de los argumentos de validez miraron a las consecuencias (CS) sólo 3 en el escenario 7 y ninguno en el 8-. Si nos posicionamos en una perspectiva progresiva de validez como la que nos propone

Messick (1989b), estos argumentos de validez (CS) son mucho más integradores de los distintos aspectos de interpretación y consecuencias que hacen a la validación de una determinada nota. Su práctica ausencia entre los comentarios -de nuestra muestra- de estos escenarios de mejora habla de la pobreza argumental y la falta de legitimación entre los docentes de una práctica que se alinee con una evaluación formativa y procesual y - como proponen muchos autores (Shepard, 2006; Ravela *et al*, 2017) la posibilidad de darle un enfoque formativo incluso a las instancias sumativas de evaluación.

Veamos algunos comentarios que nos pueden enriquecer la perspectiva:

“Se trata del segundo trimestre, en el último premio el esfuerzo de todo el año otorgando algún punto más en la calificación” (Test 6); “considero que podré ver avances en él durante el tercer trimestre, seguramente” (Test 7). En este caso, varios profesores parecen considerar la posibilidad de incluir la mejora del estudiante en la nota final de la asignatura que cierra el año junto con el tercer trimestre: “espero que siga esforzándose y superándose para llegar al 7” (Test 11).

Otros comentarios parecen considerar la mejora, pero apreciarla como muy débil aún: “si bien ha mejorado su esfuerzo, no lo hizo lo suficiente” (Test 3).

En este tipo de escenarios parece haber sido aún más normativo que en los otros el ajuste al promedio, lo que se manifiesta en sus comentarios anclados a la construcción y significado de la nota en sí (VC). Sucede, sí, igual que en otros escenarios que los docentes realizan una opción, pero, a la hora de explicar por qué lo han hecho, parecen “mitigar” la elección realizada, abrir a otras posibilidades o incluso argumentar por otra posible opción, en vez de fundar la postura elegida. A veces indicando que “igualmente” hablaré con el estudiante, “más adelante” puedo tenerlo en cuenta, “incluyo o incluiría otras instancias de evaluación” – buscando más evidencias (RU) como lo señalábamos en otros escenarios-. Otro ejemplo: “igualmente deseo aclarar que no solo tomo en cuenta las evaluaciones para la nota del trimestre” (Test 9), para el escenario nro. 7.

Finalmente compartimos otro escenario elaborado específicamente para esta instancia de investigación en la que planteamos una situación paradójica en las evidencias que el docente ha recogido a través de distintos tipos de desempeño.

Cuadro 9: Escenario 9 de situación de evidencia contradictoria.

Escenario 9

De acuerdo con los objetivos presentados en el curso de 4to año de Historia, su alumno Pedro ha mostrado un buen desempeño. Ha logrado representar cada vez con mayor claridad en la línea histórica la sucesión de los acontecimientos del siglo XX en Argentina, así como la relación sincrónica con los eventos y procesos de otras regiones del mundo. En la última lección ha podido dar cuenta de esta producción preparada en la casa de forma brillante. Por otra parte, en la prueba escrita integradora tuvo un 4/5. El promedio general del período trimestral le daría un 5/6.

Ante esta situación:

N/27	%	Opciones ofrecidas por el escenario
8	29.6	a) Ud. le deja el promedio de 5/6.
18	66.6	b) Le pone un 6/7 de promedio al mirar los últimos resultados en relación con los objetivos de la materia.
1	3.7	c) Le pone un 9 porque su última producción oral ha sido excelente.

¿Por qué elegiría esta opción?

Escenario 9	Puntaje de validez progresiva				
	VC = 1	RU = 2	IV = 3	CS = 4	Total
Deja promedio 5	3	2	2	1	8
Pone un promedio de aprobación	5	7	4	1	18
Pone un promedio más alto	---	1	---	---	1
Total	8	10	6	2	27

Moda: 2 (RU)- 37.03 %

Fuente: Elaboración propia.

Es prácticamente el único escenario en que los comentarios se han volcado más sobre aquellos aspectos que hacen a una validez fundada en la relevancia y utilidad de la

nota -que incluye tanto su interpretación como el uso concreto de esa nota para el caso de ese estudiante en particular- aunque la distribución entre los 4 tipos de argumentos ha sido bastante homogénea. “Todos los trabajos son calificados, si en la última producción levantó su rendimiento se verá reflejado en su calificación final”, (Test nº15). Parece reflejar aquella convicción de que “todo va con nota”, como nos lo comentaban a partir de sus propias investigaciones Ravela, Piccaroni y Loureiro (2017).

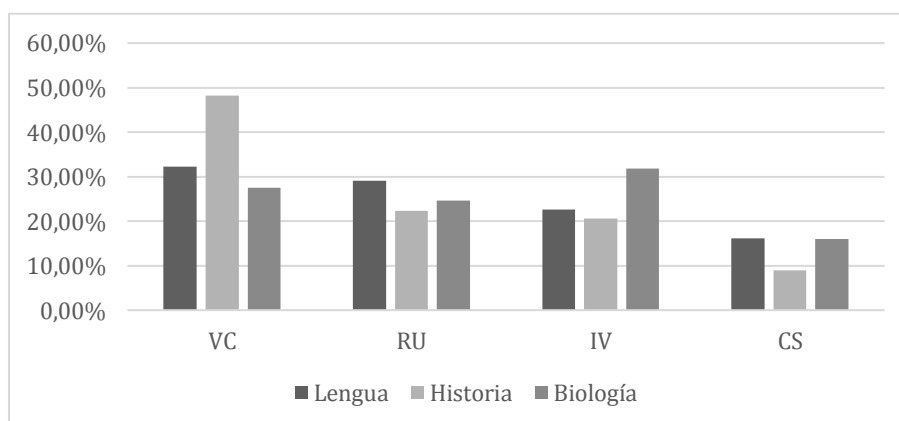
Parece que aquí perciben que han recogido una evidencia lo suficientemente clara para asignar una nota de aprobación del trimestre -lo que no les sugería la mejora en los escenarios anteriores-: “Le pongo nota 6 -aprobación en Ciudad- en reconocimiento a que logró alcanzar lo pedido y lo charlo con el alumno” (Test 16).

Un docente que propone reconocer el esfuerzo aclara que no es “tan importante lo numérico” (Test 17), marcando una incomodidad en la aplicación de la normativa reflejada en los innumerables comentarios que estuvieron acompañados de aclaraciones, propuestas de conversaciones ulteriores con el estudiante que, en su mayoría, apuntaban a que la interpretación de la calificación no fuera desalentadora de parte del estudiante.

Análisis comparativo

Vamos a cruzar algunas de las diferentes variables que forman parte de las características de la muestra obtenida de modo de indagar las tendencias que, en su comparación, eventualmente, se manifiesten.

Cuadro 10: Fundamentación de validez por asignatura



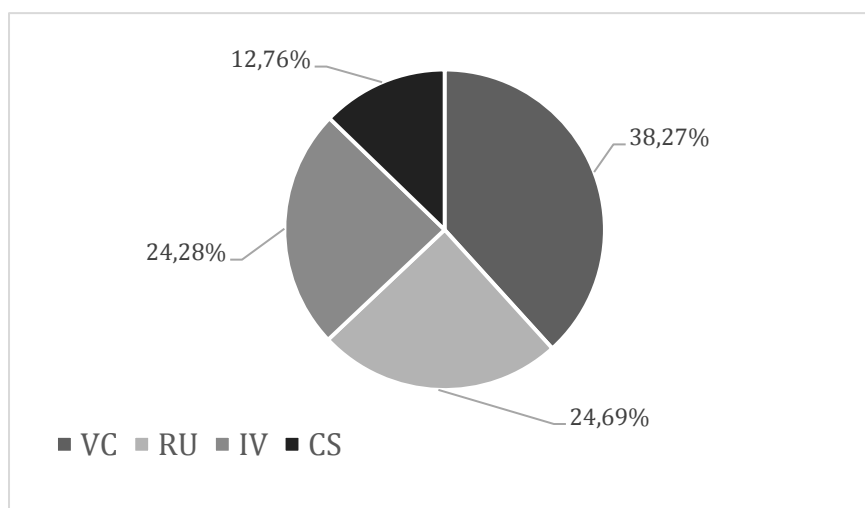
Fuente: Elaboración propia.

De la comparación entre asignaturas vemos, por un lado, que hubo siempre una tendencia mayoritaria a fundar las decisiones de calificación en argumentos de validez de construcción (VC): las pautas de evaluación y calificación establecidas e informadas por el profesor y/o la escuela que, a su vez, encarnan la normativa jurisdiccional vigente. Por otra parte, esta misma tendencia se acentúa entre los profesores de historia quienes sostienen luego un porcentaje de elección similar (en torno al 22 %) para los comentarios atribuibles a la relevancia y utilidad, y para los que guardan implicancias de valor.

Por otra parte, las explicaciones de las decisiones de calificación tomadas por los docentes de lengua y biología prácticamente se mantienen muy equilibradas entre sí, con una menor incidencias de las ligadas a las consecuencias sociales.

Si miramos Lengua y Literatura vemos menor incidencia de los argumentos más complejos de validez (CS), pero no una diferencia muy importante con los de VC, RU e IV. Y, finalmente, en Biología aparecen relativamente más frecuentemente comentarios ligados a las implicancias de valor (IV), sobre todo sobre lo justo de la actuación con cada estudiante, y bastante menos respecto de las consecuencias de valor (CS) que pudiera tener el uso e interpretación de la nota asignada.

Cuadro 11: Fundamentación de validez total por asignaturas.



Fuente: elaboración propia.

En las 3 asignaturas que hemos analizado notamos poca referencia a las consecuencias sociales de las prácticas de calificación (CS), que apenas llega a un 12,76 %. Las respuestas de los profesores de inglés de China a cada uno de los escenarios fueron

más numerosas en argumentos de validez de relevancia y utilidad (RU) y de consecuencias sociales (CS), por lo que afirman los autores que “indican que las consideraciones de los profesores tuvieron un significativo efecto en los cuatro aspectos de validez” (Sun y Cheng, 2014: 335). Observando esta tendencia desde la matriz de Messick (1989b) podemos notar en los docentes de China que influye más el uso que se le da a las evidencias y a sus consecuencias que su interpretación.

En el caso de los docentes de nuestra muestra suman entre RU y CS un poco más de un 37% por lo que el eje vertical izquierdo de la matriz que mira más a la interpretación tiene una presencia mayor entre VC e IV con un porcentaje conjunto de alrededor de un 62 %.

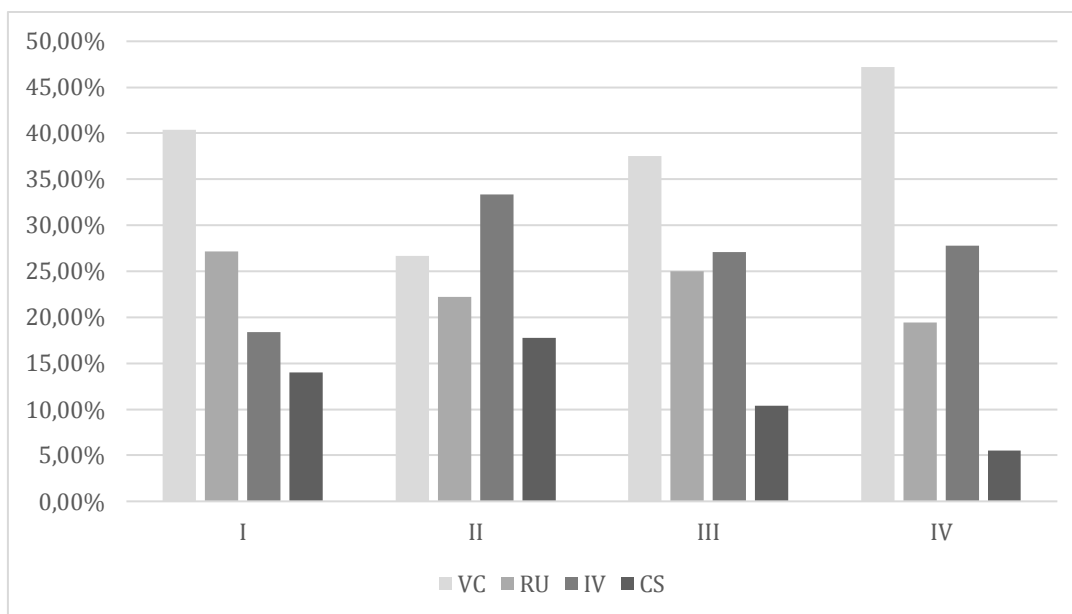
Cuando analizamos las respuestas de los docentes argentinos desde el eje horizontal de la matriz de Messick (1989b) podemos ver que hacen foco en las evidencias que sostienen las decisiones de calificación que toman en las distintas circunstancias que enfrentan. Hay una tendencia a ajustarse a la normativa y a sostener en la evidencia disponible toda decisión. La preocupación por la presión de los padres ante la disconformidad respecto de una calificación sólo está presente explícitamente en dos comentarios de una docente (Test 3), pero pareciera que las presiones del sistema que va transformándose desde una perspectiva más expulsiva que se veía en altas tasas de desgranamiento en la escuela secundaria (Rivas, 2010) hacia un modelo más inclusivo desde las políticas educativas vigentes en lugar de generar una mirada más orientada hacia las consecuencias sociales que derivan de las prácticas de calificación termina dándose una conducta propensa a atarse al promedio que indica la normativa vigente.

Otro aspecto que nos parece interesante indagar es si hay alguna tendencia en la fundamentación de las prácticas de calificación que muestre diferente criterio entre los docentes de mayor o menor antigüedad.

Hemos dividido la muestra en 4 rangos de antigüedad:

- I Inferior a 10 años
- II Entre 11 y 20 años
- III Entre 21 y 30 años
- IV Más de 30 años

Cuadro 12: Fundamentos de validez y antigüedad docente



Fuente: Elaboración propia.

Vemos una tendencia bastante homogénea. La validez de construcción (el significado de la nota en sí -de acuerdo con las pautas de calificación estipuladas con los aspectos que allí incluye el docente-) que apareció como tendencia mayoritaria en los docentes -tanto a nivel general como en el análisis por asignatura- parece atenuarse a medida que transcurre la carrera docente y luego vuelve a acentuarse muy fuertemente en el grupo de antigüedad ya próximo a la jubilación. Es difícil arriesgar una interpretación al respecto pero, desde el punto de vista profesional, parece haber mayores coincidencias entre los profesores que recién comienzan y los que se están por retirar que con aquellos que tienen mayor solidez de pertenencia y permanencia en una trayectoria consolidada y aún con perspectivas de crecimiento futuro.

Los hallazgos acerca del significado que dan a las notas entre los tres países son bastante consistentes entre sí. Vemos las siguientes similitudes:

1. Consideran la nota como un pago, una recompensa al esfuerzo realizado (Brookhart, 1993; Sun y Cheng, 2014), algo que ganan como compensación por “cierta cantidad de trabajo hecho a cierto nivel” (Brookhart, 1993: 139). “Siempre premio el esfuerzo” (Test 3; Test 17). También lo refieren en el medio latinoamericano Ravela, Piccaroni y Loureiro (2017). Esto nos llevaría a la vigencia de una concepción meritocrática que estaría fuertemente ligada a la misma

constitución de la escuela, en particular de la escuela secundaria (Sverdlick, 2012). “La nota de su trabajo refleja la calidad del mismo” (Test 16).

2. La necesidad de sintetizar en una única nota trimestral una gran cantidad de información heterogénea hace que se combinen en un único juicio aspectos totalmente diversos. Analizaremos en las conclusiones generales las distintas metáforas descriptivas que han caracterizado este fenómeno que sigue apareciendo en las investigaciones a lo largo de 100 años (Brookhart *et al*, 2016) a pesar de las recomendaciones que se vienen haciendo desde -por lo menos- la década del 90, respecto de la necesidad de calificar solo y exclusivamente con referencia a los desempeños académicos de los estudiantes (Brookhart, 1991, 1993, 2013a y b; 2019; O’Connor, 2013; Mcmillan, 2001, 2019).
3. El tema de la justicia que aparece en muchas afirmaciones de los docentes parece estar ligada al cumplimiento de lo pautado y a la equidad respecto del resto de los estudiantes de la clase -sin que esto sea una calificación normativa referida al desempeño del resto de los estudiantes-. Todos reciben lo mismo por el mismo trabajo. Especialmente porque parecen tener más un concepto ligado a la equidad que a una justicia de enfoque de derechos (a cada cual lo que se le debe).

Capítulo 5

INTERPRETACIÓN GENERAL DE LOS ESCENARIOS DE CALIFICACIÓN

A partir del análisis que hemos hecho de cada escenario, en el que hemos incluido referencias a los comentarios que reflejan los criterios de los profesores en las decisiones que toman al calificar, podemos señalar algunas características comunes a las prácticas de calificación -es conveniente recordar que no dejan de ser situaciones estereotipadas que quizás está más ligado a lo que el profesor considera que es el “deber ser” que a lo que, efectivamente realiza en situaciones similares.

Características generales de las prácticas de calificación

A partir de las respuestas de los profesores, vemos unas prácticas que muestran:

- a) La comunicación y transparencia de las modalidades, pautas y criterios de calificación transmitidos claramente al estudiante (no aparecen los padres o los colegas como necesidad de legitimar prácticas institucionales) es sumamente importante. Esto trae la contrapartida de un cumplimiento de dichos requisitos previamente avisados que no permite luego flexibilizar en función de realidades que surgen en el proceso.
- b) Una preocupación por el promedio matemático que parece no avizorar o intentar otro mecanismo que interprete mejor la norma o bien se corra de la misma. Se interpreta la misma a la letra. Ni siquiera -en su mayoría- parecen considerar la ponderación como una alternativa válida o posible.
- c) Una especie de distanciamiento y falta de responsabilidad profesional. Algunos aducen no poder intervenir en situaciones disruptivas o que demandarían – a su propio juicio- incorporar otras calificaciones o construirla de otro modo a raíz del sistema informático utilizado (Test 12). Casi pareciera que la nota surge en forma externa a la intervención del docente que dice “el promedio le dio así” (Test 4) o “si las pruebas le dieron ese promedio, no puedo subir o bajar su nota” (Test 14).
- d) Respecto de los aspectos no-académicos que incluyen en las calificaciones aparece sobre todo el esfuerzo. Hay expresiones contradictorias aún en el mismo

docente que en un lugar dice -por ejemplo- “debemos considerar el esfuerzo de la alumna y su responsabilidad” (Test 23) y, en otro escenario, “calificar es poner un número y es imposible medir el esfuerzo” (Test 23). Parece haber cierta tendencia similar -éste es un docente de Ciencias Biológicas y varios de ellos lo señalaron- a la hallada por Bailey (2012). Sin embargo, todos mencionan la relevancia del esfuerzo como parte de la nota y, por ello, incluida dentro de la validez de construcción ya que para los docentes la nota significa, digamos, resultados en pruebas, algún otro desempeño (trabajos prácticos, exposiciones orales, etc.) y esfuerzo. Entre estos factores no-académicos también mencionan “falta de compromiso” (Test 17). Eventualmente penalizan, muy pocos, la entrega tardía. Ninguno reflejó preocupación por su conducta o asistencia como determinante al momento de valorar dicho esfuerzo. Quizás sí les interesó relevar la evidencia posible (RU) que manifestara “de acuerdo con su voluntad de mejorar” (Test 2)

Estos factores que llama Bonner (2019) “*academic enablers*” (facilitadores académicos), Bowers (2009) habla de “*school success factor*” (factores de éxito escolar) o “*conative skills*” (factores conativos: interés, voluntad, compromiso, autorregulación, etc.) en la expresión de Willingham (2002) son importantes para los profesores a la hora de calificar y necesitan, indudablemente, mayor clarificación y explicitación para no oscurecer el significado de la nota.

- e) Señalaron la importancia de “acuerdos” o “arreglos” con el docente: si fue advertido, si se le “prestó un punto” (Test 2) o si el “criterio de evaluación ya estaba pautado” (Test 23), etc. -quizás debiéramos incluir estos convenios (Merle, 2007) dentro de la actitud en clase o respecto de la asignatura. Estos aspectos no aparecieron mayoritariamente, pero evidencian las restricciones de la normativa a la que prefieren ajustarse algunos docentes, aún por caminos informales
- f) La nota parece estar escindida de la enseñanza y evaluación. Sólo algún que otro comentario reflexionó sobre “los contenidos de las evaluaciones” (Test 1) en tanto estuvieran o no vinculados con otra evaluación para ponderar en la calificación la eventual mejora.

Muchos ponen ejemplos de instrumentos variados para evaluar, incluso hablan de rúbricas -en poquísimos casos-, pero parece que las calificaciones tienen una “vida independiente” y no hay ponderación de la evidencia de aprendizaje (Anijovich, 2018) que recogieron las evaluaciones a la hora de tomar decisiones ligadas a su

calificación. Sólo “cuentan” los resultados, la calificación efectivamente obtenida en una producción las que se suelen mezclar con mayor o menor claridad y nivel de explicitación al estudiante.

CAPÍTULO 6

CONCLUSIONES

PRESENTE Y FUTURO DE LAS PRÁCTICAS DE CALIFICACIÓN

Queremos plantear las reflexiones finales, conclusiones y líneas de investigación ulteriores que surgen de nuestro estudio ubicándolas claramente en el marco de la realidad de las políticas de evaluación, calificación, acreditación y promoción que el Ministerio de Nación y las dos jurisdicciones cuyos docentes hemos consultado están revisando, reformulando y comenzando a implementar.

Es en ese contexto en el que queremos sumar un aporte desde la descripción de las prácticas de calificación concretas que llevan adelante los profesores de nivel secundario y aquellos argumentos que las sustentan. Son estos docentes, con estos supuestos, creencias y criterios los actores protagónicos de este cambio. Por ello es fundamental conocerlos para explicitarlos con claridad en estas dinámicas de innovación y transformación

Regímenes de Evaluación y calificación en Argentina 2019

La educación secundaria en Argentina está en pleno proceso de cambio. Ya desde la década del 90 con la Ley Federal de Educación y otros instrumentos legales se activó una transformación con idas y vueltas siempre afectados por distintos signos políticos en la gestión del sistema, con distinto ritmo según las jurisdicciones, pero que -también con algunas tensiones en sentidos contra cíclicos como el referido a la descentralización y recentralización- ha producido cambios profundos.

La ley de educación nacional en vigencia (2006) planteó la necesidad de una revisión que no sólo atravesase el diseño curricular sino también algunas normativas fundamentales en la organización de la educación secundaria que permanecían prácticamente sin cambios desde la década del 40 del siglo pasado y por tanto, legitimadas en su intocable permanencia.

Ya en 2009, dos resoluciones del Consejo Federal comienzan a plantear la elaboración de un nuevo régimen académico. El art. 9 de la resolución 84/09 marca un claro encuadre, pidiendo:

Redefinir en un plazo máximo de tres años el régimen académico que regula formas, estrategias, criterios y momentos de la evaluación y la acreditación en el nivel, mediante procesos que garanticen la participación de directivos, docentes y alumnos, y posibiliten la asunción de compromisos y responsabilidades compartidas por parte de los distintos actores implicados (CFE Res. 84/09).

La resolución 93/09, en su art. 138 aclara aún más lo referido a las calificaciones:

Los procedimientos actualmente vigentes para calificar a los estudiantes recurren a una escala numérica para valorar la complejidad de los procesos del aprendizaje, que resulta insuficiente para dar cuenta de los mismos. La escala numérica responde, entre otras cuestiones, a una exigencia organizacional propia de la masividad de la institución escolar. Sin embargo, resulta inadecuado simplificar una decisión de vital importancia para la trayectoria escolar de los jóvenes a una cuestión de operaciones con cantidades que representan momentos particulares y no el desarrollo completo de un proceso. La calificación que el estudiante obtiene en cada período o año escolar no debe limitarse a una operación cuantitativa como es el promedio aritmético (CFE Res. 93/09).

La Provincia de Buenos Aires, luego de definir un nuevo “Reglamento General de Instituciones Educativas de la Provincia de Buenos Aires” (Decreto 2299/11) -para todos los niveles y modalidades-, establece también un nuevo “Régimen Académico Común para la Educación Secundaria”, por Resolución de la DGCE 587/11. A este régimen se le incorporó la referida a las comisiones evaluadoras adicionales (DGCE 1480/11) que ofrece una tercera instancia de evaluación en mesa de examen para una asignatura a aquellos estudiantes que, al finalizar el ciclo -luego de haber transitado también las instancias de diciembre y febrero-, adeuden hasta 3 materias. De este modo establece un mecanismo nuevo que tiende a facilitar mejores condiciones para trayectorias continuas y completas que las distintas resoluciones mencionadas establecen como eje de la política de educación secundaria. Sin embargo, no hubo grandes cambios en las normas de calificación y acreditación de las asignaturas, aunque la promoción haya experimentado una novedad con la incorporación de las comisiones adicionales.

En la Ciudad de Buenos Aires se mantenía con pequeños ajustes una normativa de calificación -la 94/92- incluso anterior a la autonomía de la jurisdicción y a la transferencia de los servicios educativos de nivel secundario. Al mismo tiempo que Provincia de Buenos Aires pone en marcha las comisiones adicionales, la Ciudad establece con su resolución 11684/11 una nueva dinámica en las instancias complementarias de enseñanza y evaluación de diciembre y febrero, centrada más en los aprendizajes prioritarios. En 2015 comienza la implementación de la Ley Educación Nacional 26.206 con la puesta en marcha de la NES -Nueva Escuela Secundaria-, sin grandes cambios en la normativa de evaluación y calificación. Un informe situacional de diagnóstico en diciembre de 2017 plantea la “eliminación de la nota numérica como única pauta de evaluación” (Equipo CABA, 2017)

En este contexto, ambas jurisdicciones, impulsadas por el Ministerio de Educación de la Nación en el marco de su política “Secundaria Federal 2030”, comienzan un proceso de revisión de diversos aspectos de la escuela secundaria, que dan en llamar “Secundaria del Futuro” -en el caso de Ciudad de Buenos Aires- y “Nuevo Formato de la Escuela Secundaria” en Provincia de Buenos Aires. En 2018 se inicia el primer ciclo de dicha experiencia pedagógica en algunas escuelas de cada jurisdicción. El sistema se refiere a ellas como escuelas promotoras.

Algunos alcances y rasgos comunes que afectan al régimen de acreditación y promoción son:

- Plantear el ciclo básico del nivel como una unidad, dividida en 4 cuatrimestres y 8 bimestres.
- Al finalizar el 1° y el 3° bimestre de cada ciclo lectivo se envía un informe con calificación conceptual-orientadora: Insuficiente, Inicial, Suficiente, Avanzado, Destacado.
- Al finalizar el primer y tercer cuatrimestre se envía un informe cuatrimestral a la familia y al estudiante con calificación numérica a partir de las evidencias y del informe. Se generan dispositivos y estrategias de acompañamiento a las trayectorias escolares.
- Al finalizar cada cuatrimestre el equipo de profesores de cada área evalúa al estudiante en forma colegiada a través de una rúbrica teniendo en cuenta el desarrollo de las ocho (8) aptitudes definidas en el diseño curricular.
- Promoción final ponderada (el original pone esta palabra en mayúsculas) que, a partir de las múltiples y variadas “evidencias” define si el alumno continúa en el período de orientación y evaluación de Diciembre / Febrero.
- Luego del período de orientación y evaluación de Febrero, se establece una nueva evaluación colegiada entre los docentes para definir con una mirada integral -en caso de ser necesario- dispositivos y estrategias de avance y permanencia.
- Con 3 o más espacios curriculares no acreditados permanece en el mismo año profundizando los saberes acreditados y recuperando los no acreditados.

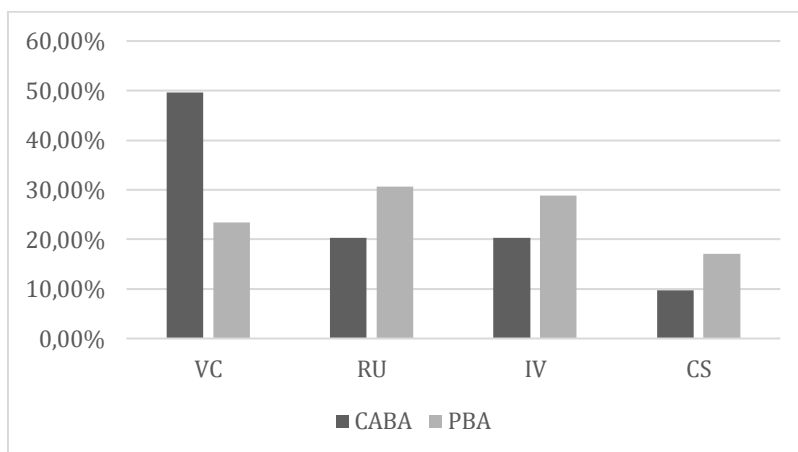
Este régimen de evaluación y calificación ofrece un encuadre teórico de evaluación formativa (Dirección de Educación Secundaria de la Provincia de Buenos Aires, 2018), fundamentada en claras evidencias de aprendizaje (Anijovich, 2017).

Sólo hemos destacado los aspectos que involucran a la calificación.

¿Son los mismos los desafíos para ambas jurisdicciones?

Veamos comparativamente cómo han utilizado los argumentos de validez los docentes de Provincia de Buenos Aires y de Ciudad que participaron del test:

Cuadro 13: Fundamentos según matriz de validez progresiva, por jurisdicción.



Fuente: Elaboración propia.

Es muy interesante que los docentes de Ciudad de Buenos Aires, cuyo nivel secundario no comenzó su transformación sino hasta el año 2015, tengan casi el 50% de su fundamentación de las decisiones de calificación en la validez de construcción (VC). Recordemos que la matriz es progresiva, por lo tanto, CS siempre expresa una más rica complejidad en los aspectos que involucra un docente al asignar una nota trimestral a un estudiante. Evidentemente han sido más completas y complejas las explicaciones de sus colegas de la Provincia de Buenos Aires.

Los profesores de esta jurisdicción, al estar sometidos a distintas transformaciones desde 1997, parecen sostener mejor la validez de sus prácticas y ser -por ende- más permeables a incorporar cambios en su forma de evaluar y calificar. En el horizonte de regímenes nuevos en marcha es un aspecto a tener en cuenta al asignar los recursos, potenciar las experiencias de innovación y estudiar tanto las prácticas hoy vigentes como la forma en que realmente vaya impactando la reformulación del régimen académico.

Algunas ideas claras para seguir pensando.

En primer lugar, queremos destacar la provisoriedad de nuestro análisis dado lo exiguo de nuestra muestra. Como hemos señalado en el estado del arte, es escasísima la

investigación sobre calificación en Argentina y América Latina, por lo que creemos que es importante haber comenzado a describir estas prácticas que son cruciales para poder garantizar los derechos que consagra nuestra Ley Nacional de Educación, que promueven una educación de calidad en trayectorias continuas y completas.

La matriz de validez progresiva de Messick (1989a) nos ha permitido dos cosas: poner el foco en la validez de las calificaciones como información y, a su vez, mirarla en toda su complejidad frente a los actores involucrados en su producción (estudiantes, profesores, directivos, escuelas) y en su uso dentro y fuera del sistema educativo (escuelas, jurisdicciones, estudios superiores y mundo del trabajo).

Podríamos inferir de los resultados y las interpretaciones compartidas, ciertas conclusiones.

1. Hay un apego a la norma escrita. Quizás sólo sea “mostrar el apego a la norma” o incluso un mecanismo consciente o inconsciente de defensa ante posibles cuestionamientos a las decisiones de calificación tomadas. Es clásica la expresión de directivos y docentes de ambas jurisdicciones: “Hay que ver que llega de inspección, qué baja el ministerio”, aunque sólo sea un tema para una jornada de capacitación, por ejemplo. Este paradigma de (no) autonomía institucional -que aquí mencionamos pero que no definiremos ni fundamentamos en este trabajo- opera también fuertemente a la hora de interpretar el régimen académico, analizar sus fortalezas, debilidades y oportunidades o, quizás, considerar un modelo de boletín diferente que incorpore otras informaciones sin dejar de incluir la que el régimen indica como obligatoria.

Es un tema a investigar e incentivar desde las políticas públicas su reflexión en los distintos estamentos del sistema. Ya recordábamos que toda calificación es en contexto, toda práctica educativa está inserta en una cultura institucional. Si todo debe estar explícitamente escrito o ser consultado a inspección, aunque se generen nuevos formatos y regímenes académicos la falta de autonomía explícita e implícita no permitirá hacer de ellos una oportunidad real para un aprendizaje significativo y una evaluación formativa.

Estas ideas surgen a partir de los comentarios de los profesores sobre la fundamentación de sus decisiones de calificación que estuvieron fuertemente orientados, en su gran mayoría a la validez de construcción y, en menor medida al uso de las evidencias reunidas para poder dar cuenta del origen de esas calificaciones ante el

estudiante y ante los padres. Como decíamos en el análisis de los escenarios, es el eje de las evidencias el que más se muestra en las preocupaciones de validación de las decisiones de los docentes al calificar. En ese sentido nos preocupa que el enfoque de “evidencias de aprendizaje” que los primeros documentos del Nuevo Formato de Escuela Secundaria de Provincia de Buenos Aires han tomado de Rebeca Anijovich (2017) sean claramente explicados en instancias de formación en servicio que permitan una reflexión sobre las propias prácticas que se quieren mejorar.

2. La validez de construcción entre los profesores testeados no ha estado orientada a los aprendizajes que consideran valiosos y cómo se recogen sus evidencias para las notas. En muchas ocasiones hablan de la importancia de comunicar cómo van a ser calificados y qué instancias de evaluación sumativa van a formar su nota trimestral. Es esto lo que el profesor considera pautas de evaluación que necesita informar y explicitar. Casi exclusivamente qué va al promedio, qué va con nota, qué notas tienen eventualmente más peso. Incluso, frente a distintos escenarios, el mismo profesor parece incluir aspectos o producciones o evidencias diferentes dentro de las pautas de evaluación que él dice haber definido con sus estudiantes. Parecen cambiar el programa de evaluación (Camilloni, 1998) sobre la marcha.

Es urgente la articulación decidida entre enseñanza, aprendizajes, evaluación y calificación. Sobre todo, la disociación de la calificación de la evaluación que hemos visto extensamente en los comentarios del profesor no parece ser un tema que preocupe y es clave porque atraviesa un dispositivo, que lo hemos dicho varias veces, es vertebrador de la escuela secundaria. El profesor parece experimentar que evalúa algo, pero la calificación le obstaculiza lo que quiere realmente transmitir

Hoy podemos acceder con cierta facilidad a las propuestas de investigadores y capacitadores como Susan Brookhart, Douglas Reeves, Ken O'Connor, Rick Wormerli, Thomas Guskey, Robert Marzano quienes vienen difundiendo principios claros de calificación en sus materiales de formación y difusión que hoy también tienen una enorme disponibilidad a través de las redes, capítulos abiertos de sus libros en internet, videos y webinars orientados a sólidas prácticas de calificación: referirlas a expectativas de aprendizaje del diseño curricular, que las notas sean una adecuada representación de los logros académicos separando los aspectos no-académicos en la información, combinar cuidadosamente los resultados de distintas producciones para que su ponderación refleje

la importancia de los aprendizajes evaluados, comunicar las notas y su significado de modo que sean adecuadamente comprendidos por estudiantes, padres y otros docentes.

Es fundamental que, desde las instancias de formación inicial y permanente - especialmente frente a un régimen académico innovador que propone un interesante “golpe de volante”- se ponga el foco en la evaluación y calificación integradas a la enseñanza.

Es muy interesante la incorporación de las rúbricas como dispositivo de evaluación para las propuestas innovadoras de la Escuela Secundaria 2030. Ofrecen un adecuado instrumento para poner el foco en las capacidades que prioriza el diseño curricular, tener una mirada integradora del estudiante en el que cada docente mire a esas aptitudes desde su asignatura, brindarle una herramienta que le indique qué ha logrado aprender, qué le falta y cómo puede orientarse en su mejora. Por otra parte, es fundamental que se establezca claramente qué entendemos por rúbricas y qué diferencia esta rúbrica “general” de las que puedan elaborar y utilizar los docentes para la evaluación y calificación cotidiana en su grupo de clase. Así mismo es muy importante que las rúbricas integren la calificación en sus niveles de desempeño descriptos ya que tienen también ese potencial. Parece un dispositivo muy poco utilizado aún en la escuela secundaria tanto de Provincia del Buenos Aires como de la Ciudad de Buenos Aires, aunque comienza a estar en el vocabulario de los docentes paulatinamente.

3. Otra preocupación que expresan con mucha frecuencia los profesores es lo que llaman una “nota justa”, “ser justo”, “no ser injusto con los demás estudiantes”. Varias investigaciones de Robin Tierney (2011; 2013 y 2015) señalan que la falta de adecuación de los docentes a prácticas de calificación más transparentes de las que, muchas veces, tienen conocimiento, obedece a que cierto conjunto de supuestos no analizados y un cierto sentido de “lo que es justo para el alumno” obturan las buenas prácticas de calificación.

Dado que la justicia o la equidad son objeto de múltiples y conflictivas interpretaciones en la literatura sobre evaluación, es fundamental un permanente diálogo entre los interesados ya que, aunque parezca extraño, el significado subyacente es frecuentemente el mismo” (Tierney, 2013: 139).

No hemos señalado cómo afectan a los profesores sus propias experiencias de ser evaluados y calificados (Brookhart, 2009; McMillan, 2019) a la hora de establecer sus criterios y clarificar los supuestos que operan en sus prácticas de calificación. Es pensarse desde la propia historia personal educativa. Nadie comienza a tener su percepción de la

nota, por ejemplo, desde que entra al profesorado o comienza a ser docente en un colegio. Sus experiencias como estudiante son fundantes y estructurantes de su práctica. Es necesario generar dentro de los espacios de capacitación (para los nuevos regímenes académicos u otros) dispositivos que permitan reflexionar sobre dichas experiencias y poder pensar cómo construyeron en los profesores, determinados supuestos de justicia, injusticia, equidad, etc. Son los fundamentos axiológicos profundos de todo modelo de enseñanza, aprendizaje, evaluación y calificación que no pueden estar ausentes de una implementación que quiera ser coherente y, eventualmente, factor de crecimiento para el sistema educativo.

4. Frente a la preeminencia de esta validez de construcción: qué incluimos o no en la nota trimestral y ajustarse al promedio como mecanismo omnipresente y soberano, los profesores que, también en muchísimos comentarios detectan diversas situaciones pedagógicas que afectan las trayectorias escolares, parecen querer compensar esas decisiones “duras”, eventualmente atadas a una lectura literal del régimen de calificación vigente. Lo hacen por distintos caminos: incluir otros actores que permitan un mejor acompañamiento del estudiante - aunque su nota ya haya sido definida del modo explicado- como el departamento de orientación, conversación con los padres, involucrar a profesores tutores o bien el diálogo cuya necesidad ha estado muy presente en todos los comentarios, aunque haya quedado por fuera de la decisión y por tanto de la validez que la sustenta.

El diálogo como una necesidad para acompañar, para poder dar cuenta de un proceso que parece encerrarse en la normativa, en la necesidad de producir evidencia. Sería muy oportuno complementar la calificación con otro tipo de descriptores narrativos (Brookhart, 2009) En el escenario que planteaba una situación de crecimiento (escenarios 4 y 8) los docentes optaron por atenerse a los resultados y al promedio entre ellos. Parece que dar cuenta de una calificación que no se ajuste al promedio es muy incómodo para el docente. Prefiere completar una decisión con el diálogo que optar por explicitar el reconocimiento de su mejora en la calificación.

Dado que la calificación tiene una finalidad de comunicación que le es esencial, si ese diálogo no es informado en el boletín que oficializa la calificación trimestral está

irrumpiendo como algo ajeno, externo, carente de formalidad frente a la importancia que la comunidad educativa en su conjunto da a la nota.

Este diálogo que el docente experimenta como muy importante requiere de otros dispositivos o la incorporación de descriptores cualitativos en el boletín que apunten a comentar, clarificar, ampliar o explicitar el sentido que la nota no logra a transmitir, es equívoco, contradictorio o sencillamente incompleto.

Parece auspiciosa, la incorporación de calificaciones conceptuales orientadoras (en los primeros bimestres de cada bloque cuatrimestral) y la rúbrica descriptiva por aptitudes en las propuestas de la Escuela Secundaria 2030 que llevan adelante las escuelas promotoras, en su segundo año de implementación. Otro de los aspectos más interesantes de los nuevos regímenes en vías de implementación es la incorporación de instancias de evaluación colegiada que van a generar un proceso permanente de reflexión sobre las propias prácticas y su progresiva mejora. Pero, si los boletines “con notas” son diferentes e los “boletines orientadores” ya sabemos cuáles son importantes y cuáles no.

De allí que se hace necesario investigar, difundir y revisar experiencias que se están desarrollando en otros países (Guskey, 2004) que permiten integrar los descriptores de avance en el aprendizaje junto con las calificaciones para que podamos ir formando criterio de cómo poner el foco en lo más importante.

5. Este enriquecimiento de las instancias de información (boletín) puede ser también una excelente ocasión para abordar el otro fenómeno encontrado en las prácticas de calificación que lleva a mezclar distinto tipo de evidencias de aprendizaje, actitud, conducta, etc. que describen los investigadores de habla inglesa con distintas metáforas, probablemente la más utilizada sea la que acuñó Brookhart en 1991: *hodgepodge grading* -revoltijo de calificación- (Brookhart, 1991,1993,2009,2019)

En la muestra analizada han expresado los profesores incluir en la construcción de la nota aspectos ligados al esfuerzo, al cumplimiento de los requisitos, al compromiso con la asignatura y a otros facilitadores académicos (Bonner,2019). La recomendación repetida por todos los especialistas (O’Connor, 2010; Tierney,2011; Brookhart,2019) es separar la información respecto de estos aspectos que -evidentemente dada su persistencia pese a todas las recomendaciones de calificar sólo en función del desempeño académico - los profesores prácticamente en todo el mundo consideran importantes para su

desempeño posterior en la misma escuela secundaria, en los estudios superiores y el mundo laboral. “La mayoría (...) informó que usualmente se focalizan en los resultados académicos al calcular las notas de sus estudiantes, pero muchos estaban preocupados por las consecuencias a largo plazo de pasar por alto estos factores no-académicos” (Tierney, 2011: 3).

Quedan así esbozadas las líneas de investigación que pareciera oportuno impulsar. Revisión de las prácticas de calificación en forma individual y con los colegas, historia personal de calificación, supuestos de valores clave para la tarea como justicia y equidad, una concepción plenamente integrada de enseñanza, evaluación y calificación, la “forma” de implementar una norma o régimen académico con autonomía y, finalmente, los dispositivos que permitan dar cuenta de toda la riqueza evaluativa de que se dispone pero que “no va con nota”, “no figura en el boletín” y -así está en la tradición educativa- por lo tanto no cuenta.

Es el horizonte que tiene por delante la academia para que sus investigaciones vayan más allá de describir las prácticas -como hemos comenzado apenas a realizar en este trabajo- para ir a los fundamentos que las sostienen y poder actuar sobre los aspectos que las hacen sumamente resistentes a cualquier cambio.

BIBLIOGRAFÍA

- Anderson, Lorin (2018), "A critique of grading: Policies, practices, and technical matters", en *Education Policy Analysis Archives*, vol. 26,nº49, pp. 1-25.
- Aitken, N. y Mc.Kean, M. (2016) "Educational renovations: Nailing down terminology in assessment", consultado el 3 de mayo de 2016 desde:
https://www.researchgate.net/publication/300333769_Educational_Renovations_Nailing_Down_Terminology_in_Assessment.
- Allen, D. (2005) "Grades as valid measures of academic achievement of classroom learning", en *The Clearing House* vol. 78, No. 5, pp. 218-223.
- Anijovich, R. (comp.) (2010) *La evaluación significativa*. Buenos Aires. Paidós.
- Anijovich, R. y González, C. (2011) *Evaluar para aprender. Conceptos e instrumentos*. Buenos Aires. Aique.
- Anijovich, R. (2017) *La evaluación como oportunidad*. Buenos Aires. Paidós.
- Antibi, André (2012) "Redonner de l'espoir a 150.000 jeunes sans qualification. La constante macabre", Conference des grandes écoles. Nantes, Mayo de 2012. pp. 23-25 consultado el 20 de junio de 2018 desde: <http://mclcm.free.fr>.
- Bachelard, Gaston (1971), *La Formación del Espíritu Científico*. Siglo XXI, Buenos Aires.
- Bailey, M. (2012), "The relationship between secondary school teacher perceptions of grading practices and secondary school teacher perceptions of student motivation", Tesis de Doctorado no publicada. University of Missouri- St. Louis, USA.
- Barbier, Jean-Marie (1983), "Pour une histoire et un sociologie des pratiques d'évaluation en formation", en *Révue Française de Pédagogie*, nº63, pp. 47-60.
- Barnes, M. (2014), "How four simple words can solve education's biggest problem", TedxUrsuline College, consultado el 30 de junio de 2019 desde:
<https://www.youtube.com/watch?v=5-NykI2jOZw>.
- Bloom, B. (1976), *Características humanas y aprendizaje escolar*, Bogotá, Voluntad.
- Bonner, S. (2013), "Validity in classroom assessment practices" en McMillan, J.H. (ed.), *Sage handbook of research on classroom assessment*. Thousand Oaks, SAGE Publications, pp. 87-106.

- Bonner, S y Chen, P. (2019), "The Composition of grades: Cognitive and non-cognitive factors", en Guskey, T. y Brookhart, S. (2019) *What we know about grading. What works, what doesn't and what's next*, Alexandria, ASCD, pp.57-83.
- Bonvecchio de Aruani, M. y Maggioni, B. (colab.) (2006), *Evaluación de los aprendizajes. Manual para docentes*. Buenos Aires. Novedades Educativas.
- Bourdieu, P. – Chamboderon, J.C.- Passeron, J.C. (2002), *El oficio del sociólogo. Presupuestos epistemológicos*. Buenos Aires. Siglo XXI editores.
- Bowers, A. (2009) "Reconsidering grades as data for decision making : More than just academic knowledge" en *Journal of Educational Administration*, vol. 47, n° 5, pp. 609-629.
- Bowers, A. (2011) "What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school", *Educational Research and Evaluation*, vol. 17, n°3, pp.141-159.
- Bowers, A. (2019), "Report Card Grades and Educational Outcomes", en Guskey, T. y Brookhart, S. (2019) *What we know about grading. What works, what doesn't and what's next*, Alexandria, ASCD.
- Brookhart, S. (1991), "Grading practices and validity", en *Educational Measurement : Issues and Practice*, vol. 10, n°1, pp. 35-36.
- Brookhart, S. (1993), "Teacher's grading practices: Meaning and Values", en *Journal of Educational Measurement*, Vol. 30, No. 2, pp. 123-142.
- Brookhart, S. (1994), "Teacher's grading: Practice and Theory", en *Applied Measurement in Education*, vol.7, n°4, pp. 279-301.
- Brookhart, S. (2009) *Grading*. Upper Saddle River, Pearson education. 2da. edición.
- Brookhart, S. (2011), "Starting the conversation about grading", en *Educational Leadership*, vol. 69, n° 3, pp.10-14.
- Brookhart, S. (2013a), "The use of teacher judgement for summative assessment in the USA", en *Assessment in Education : Principles, Policy & Practice*, vol. 20, n° 1, pp. 69-90.
- Brookhart, S. (2013b), "Grading", en McMillan, J.H. (ed.), *Sage handbook of research on classroom assessment*. Thousand Oaks, SAGE Publications pp. 257-272.
- Brookhart, S.; Guskey, T.; Bowers, A.; McMillan, J.; Smith, J.; Smith, L.; Stevens, M. y Welsh, M. (2016) "A century of grading research: Meaning and Value in the most common educational measure", en *Review of Educational Research*. Vol. 86(4), pp. 803-848.

- Brookhart, S. (2017) *How to use grading to improve learning*, Alexandria, ASCD.
- Bourdieu, P., Chamborderon, J.C., Passeron, J.C. (2002), *El oficio del sociólogo*, siglo XXI, Buenos Aires.
- Bugnard, Pierre-Philippe (2004), "La note, invention des temps modernes", n°1, pp.89-95.
- Camilloni, A. W. de (1998), *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós.
- Camilloni, Alicia W. de (1998) "Sistemas de Calificación y regímenes de promoción", en Camilloni, Alicia R.W. de; Celman, Susana; Litwin, Edith; Palou de Maté, María del Carmen (1998), *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós, pp. 133-176.
- Camilloni, A.W. de (2004), "Sobre la evaluación formativa de los aprendizajes", en *Quehacer Educativo*, vol. XIV, n° 68, pp.6-12.
- Camilloni, A.W. de (2010), "La validez de la enseñanza y la evaluación" en Anijovich, R. (comp.) *La evaluación significativa.*, Buenos Aires, Paidós, pp. 23-42.
- Camilloni, A.W. de (2014) "Las metáforas conceptuales en la construcción del discurso pedagógico", en *Revista de Educación*, Año 5, N° 7, pp. 17-32.
- Cardinet, Jean (1989), "Evaluer sans juger " en *Revue Française de Pédagogie* n° 88, pp. 41-52.
- Camussi, S. (2015) *No soy un siete*, TedxRiodelaPlataED, consultado el 30 de junio de 2019 desde: <https://www.youtube.com/watch?v=H-yXOlzZ-eo>.
- Cappelletti, G. (2011), "La evaluación por competencias", en Rebeca Anijovich (comp.), *La evaluación significativa*. Buenos Aires. Paidós, pp.
- Carlino, Florencia (1999), *La evaluación educacional. Historia, problemas y propuestas*, Buenos Aires, Aique.
- Celman, S. (1998) "¿Es posible mejorar la evaluación y transformarla en herramienta de conocimiento?", en Camilloni, Alicia R.W. de; Celman, Susana; Litwin, Edith; Palou de Maté, María del Carmen (1998), *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós, pp. 35-66.
- Celman, S. (2015), mensaje para Cristian Saint Gemain, 27 de julio.
- Cizek, G., Fitzgerald, J. y Rachor, R. (1996), "Teachers' assessment practices : preparation, isolation and the kitchen sink", en *Educational Assessment*, vol. 3, n° 2, pp.159-179.

- Cross, L. y Frary, R. (1996), "Hodgepodge grading : endorsed by students and teachers alike", consultado el 31 de julio de 2019, desde: <https://files.eric.ed.gov/fulltext/ED398262.pdf>
- Cureton, L.W. (1971), "The history of grading practices", en *Measurement in Education*, vol. 2, no 4, pp. 1-8.
- Chambat, Grégory (2010), "Une histoire de la note...et de sa contestation. N'autre école", en *Révue de la Federation des travailleurs de l'education*, n° 25.
- De Ketele, Jean-Marie (1993), "L'évaluation conjugée en paradigmes", en *Revue Française de Pédagogie* n°103, pp. 58-80.
- Del Corral, Nilda y D' Andrea, Ana María (2007) "La evaluación y la calificación en la perspectiva de estudiantes y profesores en un Instituto de Formación Docente", Universidad Nacional del Nordeste, consultado el 10 de agosto de 2018, desde: <http://www.feeye.uncu.edu.ar/web/posjornadasinve/area4/Formacion%20docente%20y%20evaluacion%20en%20la%20formacion%20docente/area4-2.htm>
- Dressel, P. (1957), "Grades: One more tilt at the windmill" en *Basic College Quarterly* (Michigan State University), 6.
- Díaz Barriga, Angel (1994) , "Una polémica en relación al examen". en *Revista Iberoamericana de Educación* , nro. 5, Mayo-Agosto 1994, pp. 161-181.
- Dilon, Cell (2018) "Grading discrepancy in global education people", en *International Journal of Social Sciences*, vol. 4, n° 2, pp.1611-1624.
- Durm, Mark (1993), "An A is not an A is not an A: A History of Grading", en *The Educational Forum*, vol. 57, Spring.
- Eloa, N.- Zanelli, N.- Oliva, A.-Toranzos, L. (2010), *La evaluación educativa. Fundamentos teóricos y orientaciones prácticas*. Buenos Aires. Aique.
- Gipps, C. (1998), *Beyond testing-Towards a theory of educational assessment*. The Falmer Press
- Guigou, J. (1992), "Retouche pour une histoire de l'évaluation", consultado el 20 de noviembre de 2012, desde <http://tempscritiques.free.fr/spip.php?article45>.
- Guskey, T. (1994), "Making the grade: What benefits the students?" en *Educational Leadership*, Vol. 52, n°2 pp. 14-20.
- Guskey, T. (2011), "Five obstacles to grading reform", en *Educational Leadership*, vol. 69, n°3, pp. 16-21.

- Guskey, T. y Brookhart, S. (2019), "Introduction", en Guskey, T. y Brookhart, S. (eds), (2019), *What we know about Grading. what works, what doesn't, and what's next*. Alexandria, ASCD. pp.1-11.
- Guskey, T. y Link, L. (2018), *Exploring the factors teachers consider in determining student's grades*. *Assessment in Education: Principles, Policy & Practice* 26:3, 303-320.
- Gvirtz, Silvina y Palamidessi, Mariano (2008), *El ABC de la tarea docente: currículum y enseñanza*. Buenos Aires. Aique.
- Hummel Wiley, C. (2011), "Profiles of teacher grading practices : Integrating teacher beliefs, course criteria, and student characteristics", Tesis de Doctorado no publicada. The University of Arizona, USA.
- Harlen, W. (2004), "A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes", Consultado el 31 de julio de de 2019 desde : http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass_rv4.pdf?ver=2006-03-02-124724-997
- Imperial, P. (2011), "Grading and reporting purposes and practices in catholic secondary schools and grades' efficacy in accurately communicating student learning", Tesis de Doctorado no publicada. University of San Francisco, USA.
- Kelly, S. "What types of students' effort are rewarded with high marks?", en *Sociology of Education*, vol. 81, n° 1, pp. 32-52
- Kohn, A. (1994), "The issue is not what but why", en *Educational Leadership*, p Octubre 1994, pp. 38-41
- Kohn, A. (1999, "From degrading to de-grading", en *High School Magazine*, vol.6, n°5, pp. 38-43.
- Kohn, A. (2011) "The case against the grades", en *Educational Leadership* , vol. 69, n°3, pp. 38-43.
- Laugier, H. y Weinberg, D. (1927), "Le Facteur subjectif dans les notes d'examen", en *L'année psychologique*. vol. 28, pp. 236-244.
- Laugier, H. y Weinberg, D. (1930), "Le Facteur subjectif dans les notes d'examen. (Deuxième contribution) ", en *L'année psychologique*. vol. 31, pp. 229-241.
- Litwin, E. (1998), " La evaluación: campo de controversias y paradojas o un nuevo lugar para la buena enseñanza" en Camilloni, Alicia R.W. de; Celman, Susana; Litwin,

- Edith; Palou de Maté, María del Carmen (1998), *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós, pp. 11-34.
- López Pastor, V. M. (2005), “La evaluación como sinónimo de calificación. Implicaciones y efectos en la educación y formación del profesorado” en *Revista Electrónica Interuniversitaria de Formación del Profesorado*, vol. 8, n°4, pp. 1-7.
- Malehorn, Hal (1994), “Ten measures better than grading”, en *Clearing House*, Vol 67, n°6.
- Maulini, Olivier (1996), “Qui a eu cette idée folle un jour d’inventer les notes à l’école? Petite histoire de l’évaluation chiffrée à l’usage de celles et ceux qui désirent s’en passer et des autres”. Consultado el 10 de noviembre de 2012, desde www.panote.org.
- Maulini, Olivier (2003), “L’école de la mesure. Rangs, notes et classements dans l’histoire de l’enseignement” en *L’Éducateur*, Mars 2003, pp. 33-37.
- Macha, M. (2006) “La evaluación en la escuela secundaria de la Ciudad de Buenos Aires: una mirada sobre qué y cómo se evalúa a través de la prueba escrita”, Tesis de Maestría en Educación. Universidad de San Andrés.
- Martin, J. (2002), “Aux origines de la « science des examens (1920-1940) ” en, *Histoire de l’éducation*, n° 94, pp. 177-199.
- Marzano, R. (2010), *Formative assessment & Standards-Based Grading*, Bloomington, Marzano Research Laboratory.
- Marzano, R. (2018), *Making classroom assessment reliable and valid*, Bloomington, Solution Tree Press.
- McMillan, J.H. (1998), “Classroom assessment and grading practices”, consultado el 31 de julio de 2019, desde: <https://eric.ed.gov/?id=ED453263>.
- McMillan, J.H. (2001), “Secondary teachers’ classroom assessment and grading practices” en *Educational Measurement : Issues and Practice*, vol. 20, n°1, pp. 20-32.
- McMillan, J.H. (2013), “Why we need research on classroom assessment” en McMillan, J.H. (ed.), *Sage handbook of research on classroom assessment*. Thousand Oaks, SAGE Publications, pp. 3-16.
- McMillan, J. H. (2019), "Surveys of Teachers’ Grading Practices and Perceptions", en Guskey, T. y Brookhart, S. (2019) *What we know about grading. What works, what doesn't and what's next*, Alexandria, ASCD, pp. 84-112.
- Merle, P. (2007), *Le notes. Secrets de fabrication*. Paris. PUF.
- Merle, P. (2014a), “La notation des élèves. Constats et interprétations”. Conferencia dada en el Collège de France, el 7 de marzo de 2014, consultado el 30 de junio de 2019,

desde: <https://www.college-de-france.fr/site/pierre-michel-menger/seminar-2014-03-07-11h00.html>.

Merle, P. (2014b), Faut-il finir avec les notes?, consultado el 20 de junio de 2019, desde : <https://laviedesidees.fr/Faut-il-en-finir-avec-les-notes.html>.

Messick, S. (1989a), “Validity” en R. L. Linn (ed.) (1989) *Educational Measurement* (3a. ed.), New York, Macmillan, pp. 13-103.

Messick, S. (1989b), *Meaning and values in test validation. The science and ethics of assessment*. Educational Researcher, 18 (2), 5-11

Messick, S. (1995), “Standards of validity and the validity of standards in performance assessment”, en *Educational Measurement : Issues and Practice*, vol. 14, n° 4, pp. 5-8.

Moll, Marita (1998) “A brief history of grading”, en *Teacher Newsmagazine* , Vol. 11, N°3.

Morejon, S. (2011), “Las representaciones en torno a la evaluación”, en *Revista. Novedades Educativas* nro. 251, Noviembre 2011, pp. 5-9.

Moss, C. (2013), “Research on classroom summative assessment”, en McMillan, J.H. (ed.), *Sage handbook of research on classroom assessment*. Thousand Oaks, SAGE Publications, pp. 235-256.

Mottier López, L. (2010), “Evaluación formativa de los aprendizajes: síntesis crítica de los trabajos francófonos”, en Anijovich, R. (comp.), *La evaluación significativa*, Buenos Aires, Paidós.

O’Connor, K. (2009), “Reforming grading practices in secondary school”, *Principal’s Research Review*, vol. 4, n° 1, pp.1-7.

O’Connor, K. (2010a), “Grades? When, why, what impact, and how?”, *Education Canada* , vol. 50, n° 2.

O’Connor, K. (2010b), “Making the grades. Ensure accuracy, meaning, consistency and support for learning”, ASCD Express, Vol. 5. Issue 3, 12 de noviembre de 2009.

O’Connor, K. (2013), *The School’s leader’s guide to Grading*, Bloomington, Solution Tree Press- Naesp.

Palou de Maté, M. (1998), “La evaluación en las prácticas docentes y la autoevaluación”, en Camilloni, Alicia R.W. de; Celman, Susana; Litwin, Edith; Palou de Maté, María del Carmen (1998), *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós, pp. 93-132.

Perrenoud, P. (1989) “L’évaluation entre hier et demain”, en *Coordination* nro. 35, pp. 3-5

- Perrenoud, P. (2008) *La evaluación de los alumnos. De la producción de excelencia a la regulación de los aprendizajes. Entre dos lógicas*. Buenos Aires. Ediciones Colihue.
- Perrenoud, P. (2009) *La construcción del éxito y del fracaso escolar, 4a ed.* Madrid. Ediciones Morata.
- Pollio, Howard (2004), “ School Grading System”, en *Gale Encyclopedia of Education*, 2d. edition, 2004.
- Pruzzo de di Pego, V. (1999), *Evaluación curricular: evaluación para el aprendizaje*. Buenos Aires. Espacio.
- Ravela, P., Piccaroni, B., Loureiro, G. (2017) *¿Cómo mejorar la evaluación en el aula? Reflexiones y propuestas de trabajo para docentes*. Grupo Magro Editores - Secretaría de Educación Pública- Instituto Nacional para la Evaluación de la Educación, Ciudad de México, México.
- Rivas, A. (2015), *América Latina después de PISA: lecciones aprendidas de la educación en 7 países 2000-2015*, Buenos Aires, Fundación CIPPEC.
- Sadler, D.R. (2009) “Grade integrity and the representation of academic achievement”, en *Studies in Higher Education*, Vol.1, no. 1.
- Santos Guerra, M.A. (1995), *La evaluación: un proceso de diálogo, comprensión y mejor*. Buenos Aires. Ediciones Aljibe.
- Scriven, M. (2013), “The foundation and future of evaluation”, en Donaldson, S. (ed.), *The future of evaluation in society. A tribute to Michael Scriven*. Information Age Publishing.
- Schneider, J. - Hutt, E. (2014), “Making the grade: a history of the A-F marking scheme”, *Journal of Curriculum Studies*, vol. 46, n°:2, pp. 201-224.
- Shepard, L. (2006), “La evaluación en el aula”, en Brennan, R. *Educational Measurement*, ACE/Praeger Westport. (Traducción para el Instituto Nacional para la Evaluación de la Educación, México).
- Shepard, L. (2013), “Foreword”, en McMillan, J.H. (ed.), *Sage handbook of research on classroom assessment*. Thousand Oaks, SAGE Publications, pp. xix-xxv.
- Starch, D. (1913), “Reliability and distribution of grades”, *Science*. 38(983), pp. 60-636.
- Strach, D. (1915), “Can the variability of marks be reduced ?”, *School & Society* vol. 2, pp. 242-243.
- Starch, D. y Elliot, E.C.(1912), “Reliability of the grading of high-school work in English”, *School Review* ,vol. 20, pp. 442-457.

- Starch, D. y Elliot, E.C. (1913a), "Reliability of grading work in Mathematics", *School Review*, vol. 21, pp. 254-259.
- Starch, D. y Elliot, E.C. (1913b), "Reliability of grading work in History", *School Review*, vol. 21, pp. 676-681.
- Stiggins, R., Friesbie, D. y Griswold, P. (1989), "Inside school grading practices : Building a research agenda", en *Educational Measurement: Issues and Practice*, vol..8, n°2, pp 5-14.
- Stiggins, Richard (1993), "Two disciplines of educational assessment", en *Measurement and Evaluation in Counseling and Development* (American Counseling Association), Vol. 26, n°1, pp. 93ss.
- Stiggins, Richard (2002), "Assessment crisis: the absence of assessment FOR learning", en *Phi Delta Kappan*, vol.83, n°10, pp. 758-765.
- Stiggins, Arter, Chappuis, J y Chappuis, S.(2004), *Classroom assessment for students learning. Doing it right, using it well* , Assessment Training Institute, Portland, Oregon.
- Suchaut, B. (2008), "La lotterie des notes au bac", consultado el 10 de marzo de 2019 desde : <https://hal.archives-ouvertes.fr/halshs-00260958/>
- Sun, Y y Cheng, L. (2014), "Teachers' grading practices: meaning and values assigned", en *Assessment in Education: Principles, Policy & Practice*, vol. 21, N°3, pp. 326-343.
- Sverdlick, I. (2012), *¿Qué hay de nuevo en evaluación educativa?: Políticas y prácticas en la evaluación de docentes y alumnos: Propuestas y experiencias de autoevaluación*. Buenos Aires. Novedades Educativas.
- Swineford, F. (1947) "Examination of the purported unreliability of teachers' marks", *Elementary School Journal*, vol. 47, n° 9, pp. 516-521
- Tierney, R. (2011) "Being fair: Teachers' Interpretations of Principles for Standards-Based Grading", en *The Educational Forum*, vol. 75, pp. 210-217.
- Weiss, J. (1996) "Évaluer plutôt que noter", en *Revue internationale d'éducation de Sèvres [En ligne]*, vol. 11, pp. 1-13
- Welsh, M. (2019), "Standards-Based Grading", en Guskey, T. y Brookhart, S. (eds.), *What we know about Grading. what works, what doesn't, and what's next*. Alexandria, ASCD, pp. 113-144.
- William, D. (2011), *Embedded formative assessment*, Bloomington (Indiana), Solution Tree Press.

Withney, D., Culligan, D. y Brooksher, P. (2004), *Truth in Grading*. New York. Universe.

Yezbeck, D. (2011), "Grading practices: Teacher's considerations of academic and non-academic factors", Tesis de Doctorado no publicada. Virginia Commonwealth University, USA.

Documentos

CABA- Secundaria del Futuro, *Esquema del régimen de evaluación, acreditación y promoción*.

CFCyE, Resolución 84/09

CFCyE, Resolución 93/09

D.G. de C. y E. de la Provincia de Buenos Aires, *Marco General del Régimen Académico para la Educación Secundaria- Resolución 587/11*.

Dirección de Programación educativa (2004)- *Régimen de Evaluación, Calificación y Promoción para el alumnado del nivel medio en establecimientos dependientes de las Direcciones de Área de Educación Media y Técnica y de Educación Artística y de la Dirección General de Educación Superior*.

Ley 26.206- Ley Nacional de Educación

Dirección de Educación Secundaria de la Provincia de Buenos Aires, *Documentos de actualización curricular. La evaluación y la actualización secundaria. La propuesta de evaluación en el Nuevo Formato de la Escuela Secundaria*.

Ministerio de Educación- Secretaría de Innovación Educativa y Planificación (2017) *Marco para la implementación de la Escuela Secundaria Federal 2030*.

Ministerio de Educación- Secretaría de Innovación Educativa y Planificación (2017), *Secundaria Federal 2030, Régimen Académico. Escenarios posibles para innovar*.

