

Hacia una problematización del acto de medición / evaluación de la calidad de la educación escolar

Carmelo Galioto Allegra
Universidad de O'Higgins
carmelo.galioto@uoh.cl

Guillermo Marini
Pontificia Universidad Católica de Chile
gmarini@uc.cl

Resumen:

Este artículo explora la complejidad imbricada en el acto de medir y evaluar en la educación escolar. Se utilizan dos perspectivas filosóficas: por un lado, se desarrollan las implicancias de una pregunta elaborada por Biesta ¿Valoramos lo que medimos o medimos lo que valoramos? Por otro lado, se recurre al tratado sobre las Categorías de Aristóteles para aclarar la noción de medida. A partir de estas lentes interpretativas, se exploran una prueba de examinación al término de la escuela secundaria, el Examen de Madurez (Italia), para dar cuenta de dicha complejidad, ilustrar algunos de los aspectos problemáticos que conllevan las mediciones educativas y argumentar que la dimensión subjetiva del estudiante y de las relaciones intersubjetivas deben estar al centro de aquello que importa 'e-valor' en educación.

Palabras clave: Medida, Cantidad, Cualidad, Validez técnica, Validez normativa

Resumo:

Este artigo explora a complexidade embutida no ato de medir e avaliar na educação escolar. Duas perspectivas filosóficas são utilizadas: por um lado, desenvolvem-se as implicações de uma questão elaborada por Biesta: Valorizamos o que medimos ou medimos o que valorizamos? Por outro lado, o tratado sobre as categorias de Aristóteles é usado para esclarecer a noção de medida. A partir dessas lentes interpretativas, um teste de exame no final do ensino médio, o Exame de Maturidade (Itália), é explorado para dar conta dessa complexidade, ilustrar alguns dos aspectos problemáticos que as medições educacionais acarretam e argumentar que a dimensão subjetiva do o aluno e as relações intersubjetivas devem estar no centro do que é importante para o 'e-valor' na educação.

Palavras chave: Medida, Quantidade, Qualidade, Validade técnica, Validade normativa

Abstract:

This article explores the complexity embedded in the act of measuring and evaluating in school education. Two philosophical perspectives are used: on the one hand, the implications of a question elaborated by Biesta are developed: Do we value what we measure or do we measure



what we value? On the other hand, the treatise on Aristotle's Categories is used to clarify the notion of measure. From these interpretive lenses, an examination test at the end of secondary school, the Maturity Exam (Italy), is explored to account for this complexity, illustrate some of the problematic aspects that educational measurements entail, and argue that the The subjective dimension of the student and the intersubjective relationships must be at the center of what is important to 'e-value' in education.

Keywords: Measure, Quantity, Quality, Technical validity, Normative validity

Introducción

El debate respecto a los modos de medir y evaluar el rendimiento y desempeño en educación, ya sea de los estudiantes, de las escuelas o de un sistema escolar en su conjunto, es uno de los temas más discutidos tanto en la literatura académica, como en las políticas educativas contemporáneas (Grek, Lawn, Lingard, & Varjo, 2009; Lewis & Lingard, 2015; Lingard, Martino, Rezai-Rashti, & Sellar, 2015; Ozga, 2012, 2013; Lingard & Sellar, 2013). Ese interés se relaciona al uso masivo de tests estandarizados en distintos sistemas educativos, la proliferación de mediciones y datos en el quehacer diario de las escuelas, y la producción de rankings ya sea de escuelas o a nivel de países, entre otros aspectos (Biesta, 2010).

La investigación en filosofía de la educación ha abordado de varias maneras la introducción de los tests estandarizados y el aumento de las mediciones en educación. Desde un primer momento, conviene aclarar lo que distingue la evaluación, del *assessment* y del *accountability*. La evaluación es un proceso que busca apreciar logros de desempeño y a la vez formular una apreciación normativa de algo; como plantea Fournier (2005), el juicio de valor es la característica distintiva de la evaluación. Un ejemplo de ello es el acto de un consejo de curso en que los profesores reunidos evalúan a cada estudiante en relación al año escolar vivido y juzgan su promoción al siguiente año. Por otra parte, el *assessment* (Lambert & Lines, 2000) es un proceso estrictamente relacionado a la respuesta de los estudiantes a una tarea escolar: exámenes, tests, ensayos; los exámenes finales de un ciclo de estudio o una prueba para acceder a la universidad son actos de *assessment*. Finalmente, el *accountability* en el ámbito educativo se refiere a la práctica política y organizacional de usar la información recogida por el proceso de *assessment* para dar cuenta del funcionamiento de un colegio o de un sistema escolar, a través de sanciones o incentivos (Figlio & Loeb, 2011): el uso de dispositivos como el SIMCE¹ en Chile es representativo de *accountability* en ámbito educativo, así como las políticas *No child left behind* o *Race to the top* en Estados Unidos.

Teniendo esto en cuenta, ya en 1999 Davis argumentaba en torno a la necesidad de un tratamiento filosófico del *assessment*, presentando la tarea del filósofo como la de un ‘terapeuta’

¹ El Sistema de Medición de Calidad Educativa (SIMCE) evalúa los resultados de aprendizaje de los establecimientos, midiendo el nivel de logro de los contenidos y habilidades del currículo vigente, en diferentes asignaturas o áreas de aprendizaje, a través de una prueba que se aplica a todos los estudiantes del país que cursan los niveles evaluados. Cfr. Página Web <http://www.agenciaeducacion.cl/evaluaciones/que-es-el-simce/>

de los conceptos usados respecto a los exámenes y evaluaciones educativas. Ahora bien, hoy nos encontramos en una fase histórica en que las políticas educativas han tendido a asociar el *assessment* al *accountability*. Al respecto existen investigaciones teórico empíricas que muestran las implicaciones complejas y desafiantes para el quehacer de las escuelas de estas políticas asociadas a sanciones o incentivos (Webb, 2005, 2006), así como para la formación profesional de los docentes (Wahl, 2017).

Otros autores han explorado los problemas éticos de las prácticas de *accountability* (Sellar, 2015a), de los tests estandarizados (Hayden, 2011, 2012), así como han desarrollado una lectura existencialista respecto a la angustia de las continuas evaluaciones (Harvey, 2010). Otras investigaciones se han enfocado en torno a los datos proporcionados por las mediciones y las implicancias de su omnipresencia en el sistema educativo (Sellar, 2015b; Selwyn, 2015; Thoutenhoofd, 2018), también respecto de la relación entre mediciones y la educación como fenómeno social y político complejo (Biesta, 2004, 2009, 2014). Finalmente, en 2015 un libro presenta el siguiente título: *Educational Assessment on Trial* (Davis & Winch, 2015). Ese es un caso ejemplar de filosofía de la educación que pasa al cedazo crítico el *assessment*, confrontando argumentos que sostienen tesis diferentes acerca del rol de las pruebas en educación.

Nuestro artículo quiere insertarse en esta discusión filosófica, colocándose en una fase reflexiva previa al de las consecuencias de las pruebas estandarizadas y de su crítica. El propósito del artículo es interrogar el acto de evaluar, junto con la noción de medida y las valoraciones asociadas a ese acto: nos ocupamos entonces de la relación entre evaluación y *assessment* en educación, junto con su pretensión de medir el logro de aprendizajes y los progresos de los estudiantes y de formular un juicio de valor acerca de ellos. Como señalábamos más arriba, el *assessment* busca evaluar, medir y documentar los progresos de los estudiantes en su educación y aprendizaje, a través de una batería de instrumentos. Por lo tanto, el *assessment* permite vincular dos acciones analíticamente distintas como son evaluar y medir.

Para llevar a cabo dicha interrogación, el primer paso será desarrollar un marco teórico constituido por dos lentes interpretativas: por un lado, las implicancias de la pregunta formulada por Biesta (2010, 2014): ¿valoramos lo que medimos o medimos lo que valoramos? Por otra parte, la presentación de Aristóteles sobre los modos de medir cantidades y cualidades,



y las posibles relaciones entre ambos en su tratado sobre las *Categorías*. Dejaremos emerger, desde una consideración conjunta de estas lentes, una relación problemática entre el acto de medir y el acto de valorar que involucra la cuestión de la validez técnica de las pruebas de evaluación educativa y la validez normativa de la tarea evaluativa en educación. Consideramos que las categorías de Aristóteles, a pesar de la distancia temporal con los argumentos de Biesta, y si bien no se ocupa directamente de educación, ayudan a precisar, matizar y complementar los argumentos implícitos en la pregunta planteada por Gert Bieta.

Como segundo paso, ofreceremos una descripción densa de un examen que marca la conclusión del recorrido escolar de los estudiantes: el Examen de Madurez del sistema escolar de Italia que sirve para certificar la conclusión del ciclo de enseñanza media. Elegimos este examen por su particular diseño que involucra distintas tipologías de pruebas y que otorga un papel relevante al profesorado, dando además voz a las inquietudes de los estudiantes.

Como tercer paso, exploraremos las implicancias que emergen desde el uso de los lentes teóricos que hemos elegido aplicadas a la prueba de evaluación que hemos escogido. Mediante estos pasos queremos articular un ejercicio de filosofía de la educación en el cual algunas prácticas de evaluación educativa permiten aclarar la consistencia de ciertas argumentaciones teóricas y, a la vez, estas últimas ayudan a interrogar el sentido de las evaluaciones en educación.

Valorar y medir la educación según Biesta

Según Biesta (2009), en la cultura de medición educativa imperante, la *validez técnica* de los instrumentos, es decir, sus parámetros psicométricos, la alineación entre lo evaluado en los tests y lo enseñado a través del currículum, la formulación y formato de las preguntas y actividades, entre otros, ha soslayado la *validez normativa* de la educación, entendiendo con ella, por ejemplo, la referencia explícita a sus sentidos y propósitos de la educación, la pertinencia y relevancia del currículum, la contribución de la educación a la sociedad. Este fenómeno ha terminado por otorgar a los resultados de muchas evaluaciones educativas – especialmente las censales y obligatorias, así como las de carácter transicional hacia la Universidad– un significado que excede su poder predictivo. Esto ha llevado a un descuido de la responsabilidad política de argumentar y debatir respecto de aquello que una sociedad considera que es lo más valioso que puede entregarse a sí misma a través de la educación de sus nuevas generaciones. Es desde este escenario que Biesta se pregunta, a modo de provocación, ‘¿Valoramos lo que medimos o medimos lo que valoramos?’ (Biesta, 2009, 2014)².

En la primera alternativa de dicha pregunta –‘¿valoramos lo que medimos?’– lo que se considera valioso aparece determinado por los resultados de la medición. Por ejemplo, cuando se pretende que un puntaje obtenido por tests estandarizados dé cuenta de la calidad de la educación impartida en cada establecimiento escolar: en otras palabras, los resultados de las mediciones parecen convertirse en el centro de las valoraciones en torno a lo que se puede esperar de la escuela.³

Aquí Biesta llama la atención sobre la excesiva consideración otorgada a los datos de las mediciones, como si representaran el límite de aquello que podemos afirmar y esperar de la escuela. El planteamiento de Biesta se basa en la discusión de Hume sobre el problema de la imposibilidad de la derivación del deber ser desde constataciones de hecho⁴. Según Hume, una argumentación como la siguiente sería falaz: el resultado del test de matemática de un curso ha dado un 80% de logro, mientras que en lenguaje el resultado ha sido de un 40%; entonces, el aprendizaje de la matemática debe ser lo más valorado y apreciado para este curso. La falacia yace en que el deber ser en general, y en este caso específico de la educación, no se puede derivar

² Biesta plantea la pregunta como una interrogación de naturaleza política para cuestionar el enfoque de las políticas educativas de *accountability*. Nosotros trataremos de explorar aristas de la pregunta que se relacionan a su dimensión normativa, ligada al acto de medir y valorar implicado en el proceso educativo.

³ Esto se vincula a la práctica de las políticas educativas de definir conceptos desde su operacionalización. Por ejemplo ¿Qué es la inteligencia? Lo que miden los test de inteligencia; ¿Qué es la calidad? Lo que miden los test de calidad.

⁴ Para una discusión más detallada de la relación entre el *is-ought problem* y el uso de datos empíricos, cfr. Spielthener (2017).

necesariamente de los datos factuales que brindan las mediciones. Esto implica, siguiendo con el ejemplo, que los resultados de matemáticas y lenguajes son incapaces de dar cuenta del nivel de formación de los docentes, de la atención de los estudiantes, del nivel de exigencia del establecimiento o del contexto socioeconómico de la escuela, entre otros factores que pueden incidir en los resultados de un curso. Por ello, si queremos evaluar y apreciar lo que deseamos de la experiencia educativa, es necesario complementar la información factual de las mediciones con una discusión de las visiones acerca de lo que es valioso en educación (Biesta, 2014), ya que si bien el dato aislado con pretensión de objetividad dice ‘algo’ -por ejemplo, matemáticas tuvo un 80% de logro-, no da cuenta de los propósitos implícitos y explícitos de la educación.

La segunda alternativa de la pregunta –‘¿medimos lo que valoramos?’– alude a un modo de concebir la medición supeditado a un acto de valoración. En otras palabras, contempla la posibilidad que aquello que se juzgue, discuta y consensue como valioso de la educación preceda a su medición. Esto no desautoriza las mediciones, pero sí las pone en la perspectiva de ser herramientas con la pretensión de decir algo sobre un fenómeno eminentemente complejo e intersubjetivo, a través de argumentos de índole psicométrica y estadística, que constituyen su validez técnica⁵.

En nuestra interpretación, los argumentos de Biesta permiten reflexionar sobre la oportunidad de desarrollar mejores evaluaciones en educación desde la dimensión que él define como ‘validez normativa’. Si estamos leyendo bien a Biesta, la consideración de la validez normativa es uno de los motivos por el cual también su segunda pregunta –‘¿medimos lo que valoramos?’– fuera formulada en primera persona plural. Pensamos que el sujeto de la pregunta es plural porque el acto de valoración en el ámbito educativo es una cuestión intersubjetiva, que conlleva una responsabilidad de toda la sociedad y especialmente de los diversos actores involucrados en la tarea de educar. La validez normativa, pues, se desprende de los valores y propósitos deseados y vinculados al proceso educativo, es decir, desde el reconocimiento y priorización de lo que es relevante y valioso en la educación para una sociedad. En base a lo anterior, la creación de conocimientos entre maestros y alumnos, la pertinencia y relevancia de los contenidos del currículum frente a las preguntas de los estudiantes y de las demandas de la sociedad, entre otros aspectos, no pueden ser recogidos y representados exclusivamente por argumentos ligados a la validez técnica de las mediciones.

En 2014, Biesta vuelve a plantear esta misma alternativa problematizadora, esta vez desde una línea argumentativa complementaria: puesto que la educación es un servicio público a gestionar con responsabilidad, se requiere la medición de sus resultados para una administración más eficiente de sus

⁵ Las discusiones sobre cómo asegurar la validez técnica de una medición es fuente de debate entre investigadores y entre decisores políticos. Este tipo de debate excede los objetivos y posibilidades de este artículo; la cuestión que planteamos es más bien, la relación y tensión entre la validez técnica y la validez normativa en educación.

recursos tanto humanos como materiales y pecuniarios⁶. El problema con este enfoque es que, a la hora de tomar decisiones, las partes involucradas en la medición –familias, estudiantes, profesores y directivos– suelen no formar parte del proceso de justificación y reasignación de los recursos; como si se los tuviera en cuenta en tanto actores del proceso, pero no completamente como sujetos de responsabilidad. En este sentido, Biesta sugiere que una opción realmente democrática respecto de la responsabilidad política en educación implica “la capacidad de tomar decisiones sobre lo que se debería ofrecer en vez de solo tener el permiso de elegir entre una oferta preestablecida” (Biesta, 2014, p. 34). Esta opción se orienta a generar una discusión pública respecto a los sentidos y propósitos de la educación, para que se dilucide su aspecto valioso. Eso implica acoger la contribución de las mediciones, pero a la vez, ejercer el derecho y deber por plantear y exigir aquellas ‘otras dimensiones’ de la educación que pueden quedar fuera de la medida establecida. Aquí no se elimina la necesidad de ‘e-valorar’, es decir, siguiendo la etimología de la palabra, de valorar hacia fuera, de explicitar el valor de algo para hacerlo público, sino que se ponen filosóficamente en tensión las convergencias y divergencias entre valorar y medir en educación.

Medir cuantificando o cualificando según Aristóteles

De acuerdo con estas primeras reflexiones, creemos que uno de los méritos de Biesta es precisamente ayudarnos a repensar las relaciones entre valorar y medir, como algo no obvio, sino filosóficamente cuestionable. Para dilucidar las cuestiones imbricadas en la relación entre el acto de medir y el acto de valorar, y vislumbrar las consecuencias respecto de la tensión entre validez técnica y validez normativa, nos parece útil traer a consideración el Tratado sobre las Categorías de Aristóteles (1952). Como plantea Thomasson (2004), las categorías son una lista de los predicados de la realidad que pretende ser exhaustiva y que, al modo de tipos heurísticos, nos ayudan a comprender la multidimensionalidad que ciertas características otorgan a los diversos sujetos (como son la cantidad, cualidad, relación, lugar, tiempo, etc.). Además, Thomasson argumenta que trabajar a partir de las *diferencias y distinciones conceptuales* que las categorías permiten realizar, es un camino útil para diagnosticar y esclarecer varios problemas y confusiones filosóficas. Desde esta aproximación, vamos a enfocarnos en las categorías de ‘cantidad’ y ‘calidad’⁷ para profundizar implicaciones ligadas al acto de medición en educación.

⁶ En palabras de Biesta se trata de la “medición comparativa... combustible del enfoque técnico de la responsabilidad orientada a la gestión” (Biesta, 2014, p. 52).

⁷ La discusión sobre todas las categorías excede a las posibilidades de este artículo; pero, teniendo en cuenta que en Occidente ofrecen posiblemente el lenguaje de base para las aproximaciones a la realidad, llama la atención la poca relevancia que la filosofía de la educación suele darle a este texto aristotélico y a las categorías como herramientas conceptuales.

Según Aristóteles (1952), “la característica más distintiva de la cantidad es que se predicán de ella la igualdad y la desigualdad” (6^a, 26). Es decir, que tanto las cantidades continuas (como la línea y el espacio) cuanto las discretas (como el número y los términos que componen el lenguaje) pueden ser iguales o desiguales entre sí. Sin embargo, la cantidad no admite diferencias de intensidad o grado, “por ejemplo, en relación al número, lo que es tres no es verdaderamente más tres que aquello que es cinco respecto de cinco” (6^a, 21). Se sigue entonces que medir una cantidad implica comparar, por ejemplo, el número de términos empleados entre dos o más oraciones juzgando si son iguales o desiguales entre sí respecto de su número, o el espacio de dos salas de clases juzgando si son iguales o desiguales entre sí respecto de los metros cuadrados de su área.

Por otra parte, según Aristóteles la característica más distintiva de la cualidad es que se expresa a través de la semejanza y desemejanza. Es decir, que tanto las virtudes (la justicia, el coraje, etc.) como las capacidades de hacer algo (jugar al fútbol, cocinar, etc.), las cualidades afectivas (la temperatura, la intensidad de un sonido, etc.) y la figura de un cuerpo (la cuadratura, el círculo, etc.) pueden ser semejantes o desemejantes entre sí. Además, las cualidades admiten variación de grado (10b, 25). Por ejemplo, se dice que una persona es más justa que otra, o que este joven juega mejor al fútbol que aquél, o que esta torta es menos dulce que aquella galleta, o que este edificio se ve más cuadrado que su vecino. Medir una cualidad implica entonces ponderar, por ejemplo, la virtud entre dos o más personas, juzgando el grado de semejanza o desemejanza que exhiben respecto de esa misma virtud.

Ahora bien, Aristóteles se percató además que la medida de la cantidad se puede extender de un modo análogo a las cualidades, correlacionando números con cosas que no son directamente numerables, como las sociedades, el coraje, una obra literaria, la percepción sensible de la realidad y, evidentemente, el logro educativo de un estudiante. Pensemos, por ejemplo, en un caso ligado a la percepción sensible como puede ser la medición del calor del propio cuerpo. La intensidad de la temperatura corporal es usualmente registrada mediante un termómetro que encierra en un tubo de vidrio a una sustancia como el mercurio, la cual se expande y contrae según el estímulo sea más caliente o frío. Cuando el mercurio alcanza la marca de 38°C (100.4°F) ésta indica sin ambigüedades que el sujeto a quien se ha aplicado el instrumento posee esa temperatura corporal –un buen médico tendrá en cuenta, además, que el termómetro funcione adecuadamente y que haya sido bien utilizado–, y permite afirmar sin lugar a error que otro sujeto con la misma marca tendrá la misma temperatura corporal, dadas las mismas circunstancias (Heath, 1955).

Sin embargo, la precisión del instrumento y su proyección análoga sobre la medida de la cualidad del calor corporal del sujeto, nada nos dicen respecto de cómo se siente dicha persona (de hecho, esa temperatura puede ser fiebre alta para un adulto, hacer agonizar a un anciano, o apenas inquietar a quien cuida a un bebé). Trasladando el ejemplo al plano educativo, sería inadecuado pretender que la validez técnica de las

evaluaciones escolares tuviera la capacidad de agotar la experiencia del estudiante que se enfrenta a dicha instancia: esto redundaría en una simplificación pobre del valor de la experiencia educativa en una letra o número.

En educación, entonces, mientras la medida de la cantidad posee un carácter ordinal y cardinal absoluto -una A es más que una B+ y el primer estudiante del ranking se encuentra trece veces mejor posicionado que el decimocuarto del mismo ranking-, cuando hablamos de la medida de la cualidad, ésta admite intensidades infinitas. Es decir, no tiene sentido pretender comparar una cualidad con alguna especie de unidad fundamental de su género, como si dijera ‘cinco grados de responsabilidad, ocho grados de creatividad, ‘el’ mejor compromiso social posible, ‘la’ inteligencia’ (Heath, 1955, p. 51). Lo que Aristóteles sí admite es que se pueda apreciar el aumento o disminución de una cualidad desde la observación de cuán arraigada e intensa es la pertenencia de esa cualidad en un sujeto. Por ejemplo, cuando vemos que se reconoce a algunos estudiantes por ser muy esforzados o creativos o comprometidos socialmente o inteligentes; pero, es importante repetirlo, esto no habilita lógicamente la transferencia de ese ‘muy esforzado’ como criterio de comparación ordinal y cardinal.

Resumiendo: en nuestra lectura de Biesta, el acto de medir en educación se ha ceñido a una comprensión de la medida preocupada fundamentalmente por su validez técnica; esta comprensión de la medida viene influenciada por los principios y modos de operar de las políticas educativas de *accountability*. Además, las tesis planteadas por Aristóteles nos ayudan a comprender que, por un lado, la noción de medida es abarcadora sea de aspectos cuantitativos (según su igualdad y desigualdad), sea de aspectos cualitativos (según su semejanza y desemejanza) y que, por otro lado, existe una diferencia conceptual sustantiva entre la medida de las cantidades unívocas y la medida de cualidades análogas. Esta diferencia conceptual, nos recuerda la legitimidad del uso de la noción de medida en fenómenos humanos y sociales; a la vez nos ayuda a tomar en cuenta la tensión entre las mediciones que tienen como unidad de medida aspectos ligados a la sola categoría de cantidad, y las valoraciones y apreciaciones que tienen como unidad de medida argumentos vinculados a intensidades cualitativas de los estudiantes. Dada la naturaleza del fenómeno educativo como fenómeno intersubjetivo aparece relevante plantear esta problematización entre el acto de medir y el de valorar, entre la validez técnica y la validez normativa. Por todo lo anterior, compartimos el llamado de Biesta para que en educación se genere una discusión sobre qué educación se desea, y añadimos la consideración de los argumentos ligados a las intensidades cualitativas de los estudiantes, para plantear argumentos normativos respecto a la educación que una sociedad quiere.

El Examen de Madurez en Italia

En este artículo usaremos las lentes interpretativas recién desarrolladas para interrogar un dispositivo de examinación: Examen de Madurez en Italia. Dicha opción permite realizar un ejercicio filosófico que puede ayudar a reconocer aspectos problemáticos en este dispositivo de evaluación educativa y avanzar argumentos de tipo normativo respecto de lo que está en juego en dichas evaluaciones. Por esas razones, comenzaremos describiendo en qué consiste esta prueba evaluativa que se realiza en el sistema escolar italiano.

Cada junio, desde el 1923, el *Esame di Stato*, conocido también como examen de madurez es una suerte de gran rito colectivo y ciudadano que recorre Italia. Lo rinden estudiantes entre 17 y 19 años como la última condición necesaria para aprobar la enseñanza media. Está compuesto por cuatro pruebas que siguen el siguiente orden: un ensayo; una prueba según la modalidad de enseñanza media (humanista, científico, lingüístico, artístico, técnico-profesional, etc.), un test semi-estructurado y una entrevista oral.

La comisión examinadora se compone de: tres profesores del último año de educación media que conocen ya a los estudiantes y generalmente compartieron con ellos los últimos tres años, más tres profesores externos a la escuela y un presidente de comisión que es director de otro colegio. El proceso de evaluación dura alrededor de 20 días de acuerdo con el número de estudiantes involucrados, y tiene lugar en la misma escuela a la que usualmente asisten los estudiantes.

El puntaje máximo que se puede lograr es de 100 puntos que se reparten entre las tres pruebas escritas y la entrevista oral. El estudiante llega al examen con un puntaje base vinculado a su promedio del último año, que varía desde 15 a 25 puntos. La comisión dispone de un máximo de 15 puntos para la evaluación de cada una de las tres pruebas escritas (por un total de 45 puntos), mientras que, para la evaluación de la entrevista, tiene un máximo de 30 puntos a disposición. La comisión es la responsable de evaluar las pruebas y atribuir el puntaje.

La primera prueba escrita es un ensayo en lengua italiana cuyo propósito es demostrar habilidades en el uso del idioma, así como de pensamiento analítico y crítico, y cuyas temáticas son elegidas por el Ministerio de Educación. Los estudiantes pueden escoger entre 4 tipologías de ensayo escrito: un análisis de un texto literario; la elaboración de un ensayo o artículo de diario sobre ámbitos

temáticos (artístico-literario; socio-económico; histórico-político; técnico-científico); un ensayo histórico; o un ensayo sobre un tema de actualidad

La segunda prueba, también dispuesta por el Ministerio de Educación, está vinculada al tipo de enseñanza media del estudiante (matemática para un liceo científico, traducción desde otro idioma si el liceo es lingüístico, resolución de un problema de contador en el instituto técnico comercial, etc.). Por ejemplo, en el liceo técnico profesional para la agricultura, la segunda prueba podría incluir la descripción del proceso de formación del queso y preguntas sobre cuáles son los procedimientos correctivos y preventivos para un vino que se enturbia. Para el desarrollo de esta prueba, las instrucciones del Ministerio invitan al estudiante a hacer referencia a un territorio del país y a una empresa agrícola de su conocimiento.

La tercera prueba es un test de respuestas múltiples y/o con preguntas de desarrollo en torno a las otras disciplinas. Todos los profesores de la comisión evaluadora que no han visto su disciplina involucrada en la primera y segunda prueba, elaboran las preguntas de esta tercera prueba.

Finalmente, la entrevista oral buscará verificar el dominio de expresión en italiano, y las capacidades críticas y argumentativas en base a los objetivos de aprendizaje establecidos para todas las asignaturas del último año de enseñanza media. La entrevista está abierta al público y pueden asistir quienes quieran (los compañeros de cursos, familiares, amigos u otras personas del colegio). Comienza con el estudiante exponiendo un itinerario de carácter interdisciplinario elaborado por él mismo durante el último año de escuela y que procede desde sus intereses. La comisión lo interroga al respecto y eventualmente hace otras preguntas ligadas a las asignaturas. Luego, la comisión muestra al estudiante sus pruebas escritas y las pone a discusión, entregando retroalimentaciones y comentarios. El coloquio se cierra con una pregunta final que tiene que ver con las decisiones post-escuela: ¿Qué pretendes hacer? ¿Por qué? ¿Seguir estudiando? ¿Y en qué facultad o recorrido formativo? ¿O tienes proyecciones laborales?

El diseño del examen de estado ha recibido muchas críticas por su carácter academicista y por eso ha vivido variadas reformas a lo largo del tiempo. No obstante, queda una prueba que representa un hito cultural para el país; justamente por este peso social que tiene, es un examen que los estudiantes abordan con mucho nerviosismo y ansiedad ya que su personalidad viene puesta a prueba como un verdadero rito de pasaje hacia la edad adulta.

Implicancias de los lentes filosóficos para interrogar el Examen de Madurez

En esta sección, interrogamos el Examen de madurez utilizando las lentes interpretativas desarrolladas al comienzo del artículo. Sin pretender ser exhaustivos en las posibles implicancias, vislumbramos cuatro temas: a. un primer tema tiene que ver con el uso que se hace de la escala de puntos como expresión de su medida; b. un segundo tema es representado por el modo en que dicha prueba concibe el saber que se elabora en la escuela; c. el tercer tema se manifiesta en la ansiedad que los estudiantes experimentan al abordar estas pruebas; d. un cuarto tema tiene que ver con el papel que juegan la subjetividad del estudiante y las relaciones intersubjetivas en el examen de madurez.

a. Tomando en cuenta el análisis que hemos propuesto de la noción de medida, cabe destacar que el Examen de Madurez recurre a una escala de puntos. Esto significa que finalmente el estudiante, o mejor dicho su rendimiento, será asociado a un puntaje: Un número será el encargado de cuantificar el trabajo del estudiante. Ahora bien, de acuerdo con lo que hemos planteado anteriormente, mientras que los números naturales enteros, por ejemplo, entre el 1 y el 10 son 8 y sólo 8, y únicamente cabe un ordenamiento primero, segundo, tercero, etc.; una misma nota, por ejemplo, 7, puede estar 'e-valuando' un logro semejante a una medalla de honor del Congreso, sinónimo de un esfuerzo notable, o estar dando cuenta de un esfuerzo sin compromiso y poco generoso, que no llegó a dar todo de sí. En esta convergencia entre las dos 'notas', pues, se aprecia la diferencia conceptual entre los significados de cantidad y cualidad planteada por Aristóteles.

Si hacemos el ejercicio de considerar estos argumentos en relación con los planteados por Biesta, emerge que la medida concebida por las políticas educativas de evaluación tiene como referencia la validez técnica que se basa en la enumeración y cuantificación, es decir, en los modos aristotélicos de la categoría cantidad aplicados a la educación; esa es también una arista de la proliferación de datos en educación explorada por otros estudios (Selwyn, 2015). Sin embargo, desde el plano de la cualidad, la medida se vincula a lo que se aprecia de algo, a su valoración en relación a un conjunto de criterios no numerables, por ejemplo, de la justificación que puede hacer el estudiante respecto de por qué eligió un itinerario interdisciplinario específico y no otro durante la entrevista final del examen de madurez. En ese sentido, si bien se recurre a una unidad de medida numérica, el examen de madurez toma opciones de examinación específicas y cargadas de sentido pedagógico y normativo para llegar a la expresión del logro mediante una medida tanto cuantificada como cualificada.

b. Esta consideración nos lleva a examinar, desde el lente de la tensión entre validez técnica y validez normativa planteado por Biesta, las maneras en que dicho examen aborda el saber: el examen de madurez se

adscribe a una concepción del saber basado en la discusión y el debate intersubjetivo, abierto a que existan varias perspectivas respecto a un tema o problema, un saber que se trata de ligar al contexto específico en que vive el estudiante.

La validez normativa del Examen de Madurez parece radicar en la puesta en común y discusión respecto de los motivos que llevaron al estudiante a argumentar de determinado modo en sus ensayos, a elegir un itinerario de síntesis y a elegir una opción laboral o educacional luego de la enseñanza media. Si bien la expresión más clara de esta conversación se da en la entrevista ubicada hacia final del Examen de Madurez, su carácter deliberativo parece inspirar -y ‘modelar’ filosóficamente- la validez técnica de toda la prueba. Estos aspectos permiten el desarrollo y la valoración-evaluación de habilidades como el pensamiento crítico, de argumentación, las capacidades analíticas y de síntesis, además de representar la oportunidad para un estudiante de ‘sacar’ la propia voz.

c. Desde el punto de vista de la experiencia vivida por los estudiantes, estos últimos viven con un alto grado de ansiedad y preocupación esta etapa que es una suerte de proceso de iniciación o inserción en el mundo adulto. Sienten que en este examen se juegan algo importante; la noche anterior a la prueba el sueño no llega fácilmente, es una “noche de lágrimas y oraciones”⁸. En este sentido, es notable que el informe de resultados PISA (OECD, 2017) acerca del *Students' Well-Being* haga hincapié sobre la influencia de los niveles de ansiedad de los estudiantes en su bienestar y rendimiento. Ese informe señala dos elementos para paliar la ansiedad: el papel de los profesores en ayudar a los estudiantes a conocerse a sí mismos respecto de cómo aprenden y enfrentan las pruebas; el otro factor es el diseño mismo de la prueba de evaluación⁹.

Ahora bien, respecto a este segundo factor, consideramos que en fenómenos humanos complejos como la educación, desconocer o desatender la ponderación cualitativa enfatizando la medida cuantitativa, favorecería una cierta disposición hacia la ansiedad. El motivo radicaría en que mientras muchos alumnos pueden escribir ensayos excelentes, sólo uno puede ser elegido ‘el mejor de la clase’ o ‘el primero del ranking’. Por ello, si bien propuestas de organización de la evaluación como el ranking de educación media podrían enriquecerse ad infinitum con otros indicadores de desempeño, de hecho, estas iniciativas ya tomaron la opción por jerarquizar cuantitativamente el rendimiento escolar. Dicha jerarquización contiene en sí el germen

⁸ Esta es una frase presente en una canción italiana *Nocte prima degli esami* (*La noche antes de los exámenes*), escrita por Antonello Venditti y dedicada a los sentimientos que se despliegan a la víspera del examen de madurez. La canción se puede escuchar subtitulada en inglés al siguiente link https://www.youtube.com/watch?v=FxYaH-fb0E8&start_radio=1&list=RDFxYaH-fb0E8

⁹ No deja de ser paradójico que ese aspecto sea señalado por la institución transnacional que propone las pruebas PISA para evaluar y comparar entre sí los sistemas educativos de distintos países. Parafraseando el título de un trabajo de Ozga (2012) *Assessing Pisa*, sería valiosa una evaluación auto-crítica respecto a la contribución a la ansiedad generada por parte de PISA.

de la ansiedad expresado en el frenesí por escalar posiciones: No es lo mismo mirar la tabla de posiciones desde arriba que ocupar los puestos intermedios o saberse en los últimos lugares con la demanda forzosa por ascender –mientras otros descienden–¹⁰. En nuestra interpretación, la medida cuantitativa en la evaluación escolar corre el riesgo de convertirse en una usina de ansiedad omnívora, donde los estímulos de logro parecen estar ubicados en alcanzar las escasas posiciones de vanguardia, disponibles para unos pocos.

d. Esto nos lleva a considerar como la pregunta de Biesta que hemos abordado: ¿valoramos lo que medimos o medimos lo que valoramos? Se refleja en el modo en que el Examen de Madurez toma en cuenta la experiencia educativa del estudiante en su dimensión tanto subjetiva como intersubjetiva. El examen de madurez, cómo hemos visto, recurre a la escritura de un ensayo y a una entrevista oral, en que la experiencia subjetiva e intersubjetiva juegan un papel importante.

Dicha opción se vislumbra también en la decisión de entregar el papel de evaluador a la subjetividad del profesor (en el caso del examen de madurez, de un equipo de profesores de distinta procedencia) y la decisión de asegurar una presunta objetividad en la medición, mediante el recurso a test que no involucran la relación profesor-estudiante. Emerge, entonces, la cuestión de la dimensión intersubjetiva de la educación: reconociendo el papel de la relación profesor-estudiante, dando ocasiones de experimentarla como una oportunidad de discutir públicamente lo que el estudiante ha elaborado con el saber que ha explorado en los años de educación media, o más bien imposibilitando la dimensión relacional.

Además, hay una implicancia en la pedagogía subyacente, entre medir un logro alcanzado desde una performance como la respuesta a un test de elección múltiple y medir el desarrollo de un estudiante que puede argumentar ya sea de forma escrita y oral, respecto de la sociedad en que vive, de una obra literaria, y también dar cuenta de lo que quiere hacer después de los años de escuela. Una diferencia pedagógica que radica en la sugerencia que planteara Montaigne hace casi 450 años de “preguntar al estudiante por el sentido de la clase en la cual acaba de participar” o, parafraseando el argumento, preguntarle por el sentido de un texto, de un hallazgo científico que ha estudiado, de un acontecimiento del mundo o de su entorno. La diferencia reside fundamentalmente en la distinta validez normativa que se le concede a la interacción profesor-saber-estudiante, así como al reconocimiento de la inteligencia del estudiante y a su subjetividad.

¹⁰ Ese aspecto lo explora Harvey (2010) en su dimensión antropológica.



Conclusiones

La interrogación que hemos conducido en torno al acto de medir/evaluar la calidad educativa desde las lentes interpretativas propuestas ha ilustrado y ejemplificado algunas tensiones presentes en el campo educativo que se desprenden de la pregunta ¿valoramos lo que medimos o medimos lo que valoramos? Además, el recurso a la explicación de la medida desde las categorías Aristotélicas de cantidad y cualidad ha mostrado que la expresión de la medida según la cantidad es distinta a la expresión según la cualidad; a la vez, esta distinta expresión depende del tipo de cosa o fenómeno al cual referimos una cierta categoría. La consideración conjunta de la tensión entre medir y valorar y entre las categorías de cantidad y calidad deja emerger la problematización del acto de evaluación y medición en educación.

Las implicancias de estos argumentos respecto del examen de madurez muestran que los modos en que la noción de medida se liga a las evaluaciones educativas es problemática. La noción de medida, usada en educación, en cuanto expresión de la cantidad, si bien puede tener su utilidad para recoger sintéticamente el resultado de un examen, requiere ser explorada desde la vertiente de la cualidad, es decir de las apreciaciones ligadas a la experiencia educativa y a lo que se busca en ella.

A la luz de la exploración filosófica del examen de madurez planteamos que las valoraciones de una sociedad respecto de la educación no pueden prescindir de la pregunta acerca de lo que está en juego en las relaciones educativas en la escuela: es como si la pregunta ¿Medimos lo que valoramos? requiera de una pregunta previa ¿qué está en juego en la educación?; esa pregunta, formulada en el lenguaje de las categorías de cantidad y calidad, se podría expresar de esta forma ¿Cuál es –o mejor dicho– cuáles son las unidades de medida adecuadas al fenómeno educativo? Como plantea Heath (1955) las mediciones requieren cumplir con ciertas condiciones, además de las exigidas en el objeto de medición, a saber se requiere el principio, la norma o dicho más simplemente, la unidad de medida.

Eso significa que formular la pregunta respecto de lo que valoramos en la educación escolar es una manera de volver a poner el foco de atención en el fenómeno educativo mismo.



En nuestra interpretación, la validez normativa en la educación escolar se desprende y se basa en una relación circular entre el estudiante, el docente y una cosa común: la exploración e inserción en el mundo a través del saber. En esta relación circular, lo que está en juego, y que da relevancia a la relación, es el despliegue de la subjetividad, particularmente respecto de la capacidad de juicio, los talentos, las aspiraciones e inquietudes de los estudiantes: hay una excedencia que ninguna medida podrá abarcar y que a su vez podría proponerse, en cierto modo, como unidades de medida de las e-valoraciones en educación.

Si el sistema educativo se propone evaluar el rendimiento de un estudiante: ¿Qué espacio de expresión personal concede a los estudiantes? ¿Qué tipo de sujeto estudiante está implícito en los exámenes escolares? Durante los años de colegio, el estudiante vive muchas cosas que aportan a su educación: convivir con otros, plantearse temas de ciudadanía, a conocer y practicar un abanico de disciplinas y otros modos de conocer, a organizarse en centros de estudiantes; pero, ¿cuánto de estas dimensiones entran en juego en los exámenes finales? Se nos muestra un desfase, una excedencia, entre el proceso educativo que es de todo el sujeto-estudiante y la parcialización de los exámenes.

Los argumentos formulados en este artículo, por lo tanto, van en la dirección de profundizar la tensión entre validez técnica y la validez normativa de las evaluaciones. Consideramos que una educación que valora preferentemente lo que medimos cuantitativamente es una educación cuyo riesgo es descuidar -o negar directamente- la posibilidad de medidas no unívocas, no ordinales, no cardinales, no sujetas al binomio igual-desigual, lo que trae como consecuencia pensar a los estudiantes de manera incompleta: como si no tuvieran una voz propia, algo nuevo que expresar, un abanico de sentimientos y capacidades a poner en juego o nuevas experiencias a proponer, compartir y realizar desde la escuela.

Obras consultadas

Aristóteles (1952). *Categorías*. Buenos Aires: Editorial Chacras.

Baird, J. A., Hopfenbeck, T. N., Newton, P., Stobart, G., Steen-Utheim, A. T. (2014). *State of field review assessment and learning*. Oxford University Centre for Educational Assessment Report OUCEA/14/2.

Biesta, G. J. (2004). Education, accountability, and the ethical demand: Can the democratic potential of accountability be regained? *Educational Theory*, 54(3), 233-250. <https://doi.org/10.1111/j.0013-2004.2004.00017.x>

Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57(1), 1-22. <https://doi.org/10.1111/j.1741-5446.2006.00241.x>

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment Evaluation and Accountability*, 21(1), 33-46. doi: 10.1007/s11092-008-9064-9

Biesta, G. (2014). ¿Medir lo que valoramos o valorar lo que medimos? Globalización, responsabilidad y la noción de propósito de la educación. *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, 51(1), 46-57.

Canny, A., & Hamilton, M. (2018). A state examination system and perpetuation of middle-class advantage: an Irish school context. *British Journal of Sociology of Education*, 39(5), 638-653. DOI: 10.1080/01425692.2017.1377599

Davis, A. (1998). The need for a philosophical treatment of assessment. *Journal of Philosophy of Education*, 32(1), 1-18. <https://doi.org/10.1111/1467-9752.00074>

Davis, A. (2006). High stakes testing and the structure of the mind: A reply to Randall Curren. *Journal of Philosophy of Education*, 40(1), 1-16. <https://doi.org/10.1111/j.1467-9752.2006.00492.x>

Davis, A. (2009). Examples as method? My attempts to understand assessment and fairness (in the Spirit of the Later Wittgenstein). *Journal of Philosophy of Education*, 43(3), 371-389, doi: <https://doi.org/10.1111/j.1467-9752.2009.00699.x>

Davis, A., & Winch, C. (2015). *Educational assessment on trial*. Bloomsbury Academic.

Fournier, D. (2011). Evaluation. Retrieved from *Encyclopedia of Evaluation*, <http://dx.doi.org/10.4135/9781412950558>

Grek, S., Lawn, M., Lingard, B., & Varjo, J. (2009). North by northwest: quality assurance and evaluation processes in European education. *Journal of Education Policy*, 24(2), 121-133, doi: 10.1080/02680930902733022

Harvey, C. W. (2010). Making hollow men. *Educational Theory*, 60(2), 189-201. <https://doi.org/10.1111/j.1741-5446.2010.00353.x>

Hayden, M. (2011). Standardized quantitative learning assessments and high stakes testing: Throwing learning down the assessment drain. *Philosophy of Education Yearbook*, 177-185.

Hayden, M. (2012). Fair is fair: Outcome assessment, constitutive luck, and teacher evaluation. *Philosophy of Education Yearbook*, 331-339.

Heath, M. (1955). Can qualities be measured? *The Thomist: A speculative Quarterly Review*, 18(1), 31-60.

Lambert, D., & Lines, D. (2000). *Understanding assessment purposes, perceptions and practice*. Routledge.

Lewis, S., & Lingard, B. (2015). The multiple effects of international large-scale assessment on education policy and research. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 621-637. doi: 10.1080/01596306.2015.1039765

Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. Routledge.

OECD (2017). *PISA 2015 Results (Volume III) Students' Well-Being*, PISA. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264273856-en>

Ozga, J. (2012). Assessing PISA. *European Educational Research Journal*, 11(2), 166-171. <https://doi.org/10.2304/eeerj.2012.11.2.166>

Ozga, J. (2013). Accountability as a policy technology: accounting for education performance in Europe. *International Review of Administrative Sciences*, 79(2), 292-309. doi: 10.1177/0020852313477763

Sellar, S. (2015a). Transparency and opacity: Levinasian reflections on accountability in Australian schooling. *Educational Philosophy and Theory*, 47(2), 118-132. <https://doi.org/10.1080/00131857.2013.793924>

Sellar, S. (2015b). A feel for numbers: affect, data and education policy. *Critical Studies in Education*, 56(1), 131-146. doi: 10.1080/17508487.2015.981198

Selwyn, N. (2015). 'There's so much data': Exploring the realities of data-based school governance. *European Educational Research Journal*, 15(1), 54-68. doi: 10.1177/1474904115602909

Spielthener, G. (2017). The is-ought problem in practical ethics. *HEC Forum*, 29(4), 277-292. <https://doi.org/10.1007/s10730-016-9318-8>

Thomasson, A. (2004). *Categories*, *Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/categories/>

Thoutenhoofd, E. D. (2018). The datafication of learning: Data technologies as reflection issue in the system of education. *Studies in Philosophy and Education*, 37(5), 433-439. <https://doi.org/10.1007/s11217-017-9584-1>

Wahl, R. (2017). What can be known and how people grow: The philosophical stakes of the assessment debate. *Studies in Philosophy and Education*, 36(5), 499-515. <https://doi.org/10.1007/s11217-016-9525-4>

Webb, P. T. (2005). The anatomy of accountability. *Journal of Education Policy*, 20(2), 189-208. doi: 10.1080/0268093052000341395

Webb, P. T. (2006). The choreography of accountability. *Journal of Education Policy*, 21(2), 201-214. doi: 10.1080/02680930500500450

White, J. (1999). Thinking about assessment. *Journal of Philosophy of Education*, 32(3), 201-211. <https://doi.org/10.1111/1467-9752.00131>