

Aspectos técnicos a ser considerados para el análisis de un dispositivo de evaluación.

Mariana Sarni¹

Resumen:

El presente trabajo responde a la necesidad de revisar aspectos técnicos incluidos en las propuestas de evaluación que elegimos o construimos para valorar el aprendizaje de nuestros estudiantes -como los de evaluación en general-, tales como la confiabilidad, la validez, la practicidad y la utilidad que estos encierran. Pensamos que retomar estos conceptos, puede llevar a revisar la evaluación que se realiza, cualquiera sea la dimensión o ámbito del que se trate.

Palabras claves:

Evaluación, evaluación de los aprendizajes, aspectos técnicos: confiabilidad, validez, practicidad y utilidad

Justificación.

Trataremos en este trabajo algunos requisitos técnicos específicos que el docente debe tener en cuenta al momento de construir sus dispositivos de evaluación, en este caso referidos al aprendizaje de sus estudiantes.

Nos parece importante retomar estos temas ya que el profesor de educación física en nuestro caso, y cualquiera de los profesionales de la educación en general, no siempre realizan este tipo de análisis al momento de pensar sus propuestas de evaluación y considerar los datos obtenidos para sus posteriores decisiones tanto para la enseñanza como para el aprendizaje.

¹ Profesora de Educación Física egresada del ISEF. Licenciada en Ciencias de la Educación. Coordinadora Nacional del Área de las Ciencias de la Educación del Instituto Superior de Educación Física de Uruguay. Profesora de las asignaturas Didáctica y Evaluación en el Curso de Profesores del ISEF de Montevideo.

El interés entonces, es aportar una síntesis de los mismos, incluso sometidos a recorte, esto es, a la selección de aquellos que pueden aportar a nuestro ámbito y a nuestras evaluaciones, a nuestro criterio, tal vez más que otros.

El presente artículo, se engarza con el propuesto en nuestro caso, al momento de la primera edición de la revista ISEF Digital.

Tomaremos para realizar el tratamiento teórico seguido, referencias conceptuales trabajadas por Alicia W. de Camillioni (2001) complementándolas con algunos aportes de Norman Gronlund (1973). Su consideración, nos permitirán valorar la calidad técnica de la evaluación realizada.

Aspectos técnicos concretos que proponemos para ser considerados.

Se pueden dividir en cuatro: *confiabilidad, validez, practicidad y utilidad.*

1. La validez de la prueba.

Podríamos decir que hablar de validez -centrándonos en la predictiva que posteriormente explicaremos- es valorar un aprendizaje de “otro”, considerando diferentes tipos de decisiones, ya sean inmediatas o mediatas, *sujetas a un futuro²*,

² El docente que evalúa, de concebir al aprendizaje centrado en lo diverso, en lo complejo, en lo que no se ve, en lo subjetivamente significativo sobre lo socialmente significativo, debe redoblar esfuerzos para establecer en sus mecanismos de evaluación, valoraciones y posteriores toma de decisiones profundamente éticas. No olvidemos que las categorías que históricamente han sido construidas en lo social, al respecto de lo que es o no es un buen o mal alumno, poco tiene que ver al menos para la opinión común, con cuestiones de validez o confiabilidad del dispositivo, sino más bien con categorías numéricas, sean éstas calificaciones o notas encubiertas por conceptos como malo, aceptable o buen alumno y sus variadísimas escalas. Esas, hablan por la enseñanza y por el aprendizaje. Esas son las aceptadas y sustentadas por el sistema así como dicen y hacen del/al sujeto, a alguien capaz de continuar, repetir, egresar, distinguirse como mejor o peor que otro. Al mismo tiempo, valora pues cuantitativamente en la opinión del colectivo, algo que es profundamente subjetivo, como el aprendizaje de otro sujeto.

Ahora bien, tal responsabilidad, requisito del sistema dentro del campo educativo institucionalizado, debe en todo caso y necesariamente, anclarse en datos surgidos de un dispositivo de evaluación que “(...) *evalúa lo que se pretende evaluar con él (...)*” (Camilloni, 2001:76). A ese concepto se asocia la validez del dispositivo de evaluación.

Gronlund agrega: “(...) *se trata de validez, si se habla del punto hasta el cual sirven los **resultados** de un procedimiento de evaluación en aquellos usos particulares para los cuales precisamente se obtuvieron. Si los resultados han de usarse para describir los logros del alumnado, nos gustaría que representaran el logro específico que deseamos describir, que represente todos los aspectos del logro que deseamos describir y que represente eso, nada más y nada menos (...)*” (1973:85).

Dicho autor, centra entonces el análisis de la validez, en relación a la *interpretación de los resultados obtenidos*, no a la construcción de la prueba en sí.

Plantea que “(...) *el deseo natural de cada uno de nosotros, es que cualquiera que sea la prueba de aprovechamiento que construyamos o que seleccionemos, arroje resultados que sean representativos de los tópicos y comportamientos que deseamos medir (...)*” (1973:88). Ellos son para él, contenidos del programa y objetivos de la instrucción: “(...) *Al seleccionar o construir un instrumento para la evaluación, la pregunta más importante es; ¿Hasta qué punto servirán los resultados en los usos particulares para los cuales se les destinó? He ahí la esencia misma de la validez (...)*” (Gronlund, 1973: 85). Plantea que siempre es una cuestión de grados, así como aclara que son útiles para el caso particular del que se trata.

Dentro de la validez, no todos los autores coinciden en cantidad y conceptualización. Seguramente esto se deba a momentos históricos de producción de conocimiento, a sus avances en el campo de las ciencias sociales,

y a la particular lectura que cada autor propone de esos contextos temporales, sobre todo con referencia a ciertas relaciones causales entre enseñar y aprender que al día de hoy vienen siendo superadas con modelos alternativos como lo es la *transposición didáctica*³ por ejemplo, o la relectura que vienen haciendo de ella, autores propios del campo de la didáctica, incluso posiciones recientes del psicoanálisis.

Nos permitimos manifestar aquí, nuestra creencia acerca de la posibilidad de realizar un estudio de validez, relacionada al sentido otorgado al dispositivo por el evaluador, su particular construcción y su propuesta de organización del trabajo, es decir, de construcción de la prueba, previo a su aplicación. Pensamos que los resultados son la materialización numérica de lo antedicho, y de su tratamiento político por parte del docente⁴. Según Perrenoud, “(...) *Para juzgar la validez, haría falta saber con exactitud qué competencias se pretenden medir. Pero no todos los maestros sienten la necesidad de aclarar y de nombrar las competencias que tratan de evaluar en una prueba concreta (...)*” (1996:142)

Para Camilloni, son siete los tipos de validez: de contenido, predictiva, de construcción, de convergencia, manifiesta, de significado⁵, de retroacción. Gronlund, por su parte, la resume en tres tipos: de contenido, ligada a criterio y de construcción.

En nuestro caso, desarrollaremos de todos los mencionados, aquellos que a nuestro criterio permiten enfatizar el análisis de las evaluaciones en educación física y en relación con su construcción.

³ Término trabajado por Yves Chevallard (1991)

⁴ Hacemos esta salvedad, con el interés de (re) explicitar - en una suerte de objetivación relativa - nuestros propios significados en torno a la evaluación, así como al descreimiento acerca de la causalidad entre enseñanza y aprendizaje propia de los modelos didácticos técnicos, que darían lugar a establecer una buena o mala prueba únicamente en términos lo observable de lo esperado, y su traducción directa a una nota o calificación, valorada entonces como *el resultado*.

⁵ Este tipo de validez, es según Camilloni (Pág.84), un tratamiento de los autores Dintel, Herman y Knuth. En nuestro caso, trataremos el tema en relación a lo que plantea la autora, por no haber podido hacernos del material original.

En función de lo anterior, hemos seleccionado en especial, las siguientes: validez de contenido, predictiva (específicamente en torno a lo extraescolar) y de construcción. Para el tratamiento solvente de la validez de significado, deberíamos prever la obtención de *la voz de los implicados*, cuestión que se debe tener en cuenta. Sin embargo, sí consideraremos de ella lo relativo a deber proveer información al estudiante sobre los progresos en su aprendizaje.

a) *Validez de contenido.*

Para valorar la validez de contenido, tendremos que considerar la construcción de la prueba en relación a los contenidos del curso y expectativas sobre el aprendizaje, de manera tal que los resultados sujetos a análisis sean en esa línea representativos. Camilloni plantea, que una prueba tiene validez de contenido “(...) *cuando representa una muestra significativa del universo de contenido cubierto por el curso o la unidad didáctica, la clase o una serie de clases (...)* Se trata, por supuesto, en todos los casos de contenidos importantes por su significación en el aprendizaje de los alumnos (...)” (2001:79).

Compartimos con la autora, que la dificultad que debe superar es la variedad de contenidos significativos, con la brevedad exigida en los tiempos destinados a la evaluación. Dicha brevedad deberá estar presente para no desvirtuar el significado cognitivo de las respuestas.

Aporta Gronlund, la importancia de tener en cuenta algunos pasos al momento de pensar la prueba en esta clave: primero establecer por separado un listado de los tópicos principales de los contenidos a valorar y de las expectativas en las respuestas a emitir. Segundo, ponderar con la dificultad que ello apareja, la importancia relativa de cada valoración, lo cual aclara tendrá que ver con el criterio del docente, con el Centro, y otros. Tercero, construir una tabla de especificaciones aclaratorias. Finalmente, se selecciona o se construye la prueba de aprovechamiento, de acuerdo a las especificaciones realizadas. Agrega: “(...)

mientras más correspondan las partes de la prueba a las especificaciones que se indican en la tabla, mayor será la probabilidad de que las respuestas de los alumnos a la prueba tengan un grado elevado de validez de contenido (...)” (1973:89).

La validez de contenido según Camilloni, es también llamada validez curricular, *“(...) porque el criterio externo que se emplea para garantizar esta clase de validez es el diseño curricular, ya que en este se expresan los propósitos y los niveles de aprendizaje, las secuencias y las sugerencias sobre estrategias e incluso, en ocasiones, sobre actividades de aprendizaje y modalidades de evaluación (...)*” (2001:79). Por el criterio no universal de la enseñanza - por docentes, regiones, instituciones, etc. -, la validez está limitada por sus alcances.

b) Validez predictiva.

Este tipo de validez proyecta los resultados obtenidos, con el posible desempeño posterior del estudiante en relación con los contenidos evaluados. En nuestro caso, nos interesa detenernos en el interés de la prueba por predecir tanto aprendizajes posteriores como desempeños ocupacionales o su actuación social.

Un docente que tiene en cuenta la validez predictiva en la construcción de su prueba, explica Camilloni (2001), pretenderá que los aprendizajes promovidos por ella *trasciendan el espacio y el tiempo escolar*. Ello es interés general de la educación, más allá de la teoría que se sustente en relación con evaluación específicamente. Por otra parte, *“(...) el seguimiento de los alumnos para comprobar la capacidad del instrumento para efectuar predicciones adecuadas, es una operación de medición relativamente sencilla, al alcance de todos los docentes que deben encararla como parte constituyente de sus programa de evaluación (...)*” (2001:81)

Dentro de este tipo de validez, debe estar previsto que la predicción de un aprendizaje puede tener que ver con plazos poco posibles de ser definidos. Es decir, los tiempos de aprendizaje de quien aprende, no siempre están engarzados con los tiempos institucionales de enseñanza, previstos por el sistema escolar o por el propio docente, en un plazo preestablecido.

Por ello, es necesario intentar diferenciar aquello que es posible un estudiante aprenda en un corto plazo, de lo que a priori supondría un aprendizaje a realizarse en un tiempo más prolongado.

c) Validez de construcción

Tiene que ver con la coherencia de las teorías sobre enseñar, aprender y evaluar del docente en sentido estricto, tanto en un sentido amplio, del proyecto pedagógico de la Institución en sí, así como sus relaciones. Según Camilloni *“(…) es éste un aspecto que debe ser particularmente cuidado por los docentes, porque si la evaluación no tiene un fundamento teórico válido, las decisiones que se tomen a partir de la información recogida por esa vía y de las interpretaciones que se hagan de ella carecerán de legitimidad (...)”* (2001:82).

Compartimos nuevamente con la autora, su importancia, debido a que este tipo de validez condiciona a los otros tipos.

d) Validez de significado.

Ella centra su valoración en el mayor o menor significado que una prueba de evaluación tiene para quien aprende: el evaluado. Dicha validez entonces, se ubica en el estudiante y su lectura particular del dispositivo empleado por el docente, en términos de significado para sus aprendizajes.

Ahora bien, en nuestro caso cobra importancia en relación a las posibilidades que el docente otorga al estudiante, para involucrarse en el tratamiento y comprensión del contenido durante la propuesta de evaluación.

En otras palabras, cómo el docente, la prueba de evaluación que este construye, y el procedimiento seleccionado/construido para llevarla a cabo, permiten la participación del estudiante en torno a ella. De esa forma, estarían presentes docentes y estudiantes habilitados por el mecanismo de evaluación, considerando el significado de la prueba, tanto el pretendido por el docente, como el verdaderamente significativo para quien aprende. Es decir, lo que no se puede ver, simplemente por encontrarse en territorios que obedecen a una construcción subjetiva, se realizaría desde la habilitación de una propuesta dialógica, por ello intersubjetiva.

Con relación a los efectos de la evaluación en el aprendizaje de los evaluados, plantea Camilloni: *“(...) entre los efectos deseados se encuentra el de proveer información al alumno sobre los progresos y los obstáculos que encuentra en el proceso de su aprendizaje y sobre los errores que deben superar, dándole indicios claros acerca de cual puede ser el origen de sus dificultades (...)”* (2001: 84).

Por otra parte, señala que la evaluación puede incentivar al estudiante a aprender, siempre que la expectativa de la solicitud, se ubique en el intersticio entre lo fácil de superar y lo difícil de alcanzar, es decir, promueva el *desafío de aprender*. Evidentemente, no todos tenemos los mismos desafíos ni la misma forma de acceder a ellos, por lo cual mucho hay en la prueba, pero también mucho en el interés generado en la enseñanza, así como el significado subjetivo de quien aprende, y su responsabilidad para aprender. Claro que esto es tema de otro trabajo.

2. La confiabilidad de la prueba

La confiabilidad tiene que ver con la consistencia de **los resultados** obtenidos en la administración de una prueba.

Gronlund y Camilloni coinciden en afirmar que una prueba goza de ser confiable cuando vuelta a administrarse arroja similares resultados, lo que lleva a concluir que ellos tienen un alto grado de confiabilidad.

Camilloni agrega “(...) *Para que los resultados (...) puedan ser considerados dignos de confianza, deben ser estables, permaneciendo semejantes en todas las ocasiones en que se administre ese instrumento u otro similar⁶ (...)*” (2001:85). Se habla, en este caso, de estabilidad como una de las condiciones de las que dependen la mayor o menor confiabilidad. Plantea asimismo, que otras tres condiciones forman parte de ella: su exactitud, su sensibilidad y su objetividad. La exactitud, implica que lo que se pretenda evaluar esté presente en la evaluación y la sensibilidad sugiere la capacidad para detectar el grado de variaciones en las respuestas obtenidas. La objetividad supone que la evaluación es independiente de la persona que evalúa. En todos los casos, estas condiciones se establecen con más confiabilidad, cuanto más cercanos sean los dispositivos empleados a la medición.

En nuestra profesión, los test de capacidades condicionales por ejemplo, sujetos todos ellos a instrumentos de medición, serán desde esa perspectiva más confiables, ya que se valoran por instrumentos tales como cronómetro, pulsómetro, centímetro, etc.

Ahora bien, el interés en ese caso, es medir la capacidad en sí, y no la comprensión de la capacidad en relación a sus conceptualizaciones y/o a su

⁶ El subrayado en esta ocasión, corre por parte de quien suscribe, tratando de resaltar el matiz entre los autores estudiados.

repercusión de las mismas para lo que viene siendo llamado “calidad de vida”. Para esto último, otro será el tema de evaluación, y por ello, otro el dispositivo y otra la valoración de cada una de las condiciones para dar cuenta de mayor o menor confiabilidad.

Asimismo, ambas formas esbozadas pueden analizarse como válidas, aunque no necesariamente ambas son confiables. “(...) en cada caso en consecuencia, habrá de determinarse cuáles son los grados de confiabilidad que resultan necesarios para cumplir con los propósitos de la prueba (...)” (Camilloni, 2001:87/88). Entonces, “(...) cabe hacer notar que la confiabilidad por sí sola no da seguridad alguna de que los resultados de la evaluación suministren información veraz (...)” (Grondlund, 1973:86).

Por otra parte, y según Camilloni, la validez y la confiabilidad de un dispositivo de evaluación nunca son absolutas; las pruebas de evaluación llamadas “objetivas”, gozan de mayor confiabilidad -de estar bien tratadas- y pueden o no gozar de validez.

3. La practicidad de la prueba

En este sentido, entiende Camilloni, por practicidad “(...) la conjunción en una prueba de tres aspectos: su administrabilidad, la facilidad de análisis e interpretación de sus resultados y elaboración de conclusiones y, por último, de la evaluación de la economía de tiempo, esfuerzo y costos de realización. (...)” (2001:89).

Los tres aspectos deben considerarse en conjunto para valorar la practicidad, aunque se puede advertir a priori, que a menor tiempo necesario de todo lo antedicho, mayor practicidad. La practicidad es relevante al momento de permitir

realizar evaluaciones más frecuentes, con sus consecuentes beneficios para la enseñanza y el aprendizaje.

4. La utilidad de la prueba

Refiere a satisfacer las necesidades específicas de los procesos de aprendizaje y de enseñanza, y según Camilloni responde a la pregunta *¿qué usos se dará a los resultados obtenidos?* Deben ser útiles para la enseñanza del docente, así como para (re)orientar a los aprendizajes de los estudiantes o profundizarlos, para el conocimiento del docente sobre el tipo de aprendizaje del estudiante. A otro nivel, deben colaborar con la Institución, las familias, los sectores sociales interesados por la calidad del aprendizaje en el área específica.

Hasta aquí, hemos acercado algunos aspectos que nos permitirán analizar la un dispositivo de evaluación elegido/construido, de modo general, y específicamente en relación con los asuntos técnicos.

Mariana Sarni

Bibliografía

R. W. de CARMILLONI, Alicia; CELMAN, Susana; MATÉ, Carmen y LITWIN, Edith. (1998): ***La evaluación de los aprendizajes en el debate didáctico contemporáneo.*** Paidós, Buenos Aires.

GRONLUND, Norman (1973): ***Medición y evaluación de la enseñanza.*** Pax, México.