

COLECCIÓN
TEMAS DE ENSEÑANZA

La evaluación en la educación superior: un escenario de controversia

UNIVERSIDAD DE LA REPÚBLICA . COMISIÓN SECTORIAL DE ENSEÑANZA

#2



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Red de Unidades de
Apoyo a la Enseñanza



La evaluación en la educación superior: un escenario de controversia

UNIVERSIDAD DE LA REPÚBLICA . COMISIÓN SECTORIAL DE ENSEÑANZA

Universidad de la República. Comisión Sectorial de Enseñanza. Unidad Académica
La evaluación en la educación superior: un escenario de controversia.
Montevideo: Udelar. CSE, 2015 (Temas de enseñanza, n.º 2)
144 p.
ISBN 978-9974-0-1214-1

Comisión Sectorial de Enseñanza
José Enrique Rodó 1854
Montevideo-Uruguay
www.cse.edu.uy
publicaciones@cse.edu.uy

Rector de la Universidad de la República

Dr. Roberto Markarian

Pro-Rector

Fernando Peláez

Comisión Sectorial de Enseñanza

ÁREA DE TECNOLOGÍAS Y CIENCIAS DE LA NATURALEZA Y EL HÁBITAT

Titular

Cristina Friss

Alternos

Pablo Kelbauskas
Alejandro Amaya
Gustavo Marisquirena
Virginia Villalba
Carlos Luna

ÁREA SOCIAL Y ARTÍSTICA

Titular

Virginia Orlando

Alternos

Luis Senatore
Adriana Santos Melgarejo
Gonzalo Vicci
María del Luján Peppe
Mariela Gandolfo

ÁREA SALUD

Titular

María Josefina Verde

Alternos

Tabaré Ravecca

ORDEN DOCENTE

Titular

Fabiana de León

Alternos

Magalí Pastorino

ORDEN EGRESADOS

Titular

Patricia Manzoni

ORDEN ESTUDIANTIL

Titular

Mauro Font

Alternos

Lucía Yarzabal
Fabrizio Rodríguez

Unidad Académica

Mercedes Collazo
Beatriz Diconca
Nancy Peré
Rafael Rey
Sylvia De Bellis
Patricia Perera
Gabriela Pérez Caviglia
Vanessa Sanguinetti
Esther Sánchez

EQUIPO EDITOR

Coordinación

Beatriz Diconca

Diseño

Gabriela Pérez Caviglia

Diagramación

Soledad Menéndez

Tabla de contenidos

Presentación por <i>Mercedes Collazo</i>	5
Primera parte	9
JORNADA: Dilemas, conflictos e interrogantes sobre la evaluación de los aprendizajes en la Universidad	
Mesa de apertura	
• Luis Calegari.....	9
• Carolina Cabrera.....	10
• Laura Coitiño.....	12
CONFERENCIA: Nudos de debate sobre la evaluación de los aprendizajes en la Universidad, necesidades de actualización, profundización e investigación educativa por <i>Alicia Camilloni</i>	15
Debate.....	39
Segunda parte	57
JORNADA: Las pruebas diagnósticas a debate	
Mesa redonda	
• Alicia Camilloni.....	57
• Andrés Peri.....	66
• Marina Míguez.....	80
Debate.....	90
Tercera parte	105
JORNADA: ¿De qué evaluación formativa hablamos?	
Presentación y debates sobre experiencias de evaluación formativa en la Universidad de la República	
CONFERENCIA: Debates pedagógicos sobre evaluación formativa en el nivel universitario por <i>Rebeca Anijovich</i>	105
Resúmenes de las Presentaciones	
• Evaluación formativa en un curso de Histología y Biología del desarrollo mediante plataforma EVA en Facultad de Veterinaria, Udelar por <i>G. Pedrana, C. Borlido, F. Alcaide, M. Viotti, V. Riaño, P. Genovese, P. Lombide, J. Calvo, E. Souza, J. Passarini, A. Bielli</i>	115

- Evaluación Formativa en un teórico de Física.
Facultad de Ingeniería, Udelar por *A. Auyuanet, L. Amy, N. Barrios*117
- Experiencia de evaluación formativa en la Facultad de Ciencias Sociales por *L. Bevilacqua, I. Martínez, D. Pérez*118
- La evaluación formativa: una práctica compleja. INET-CFE
por *A. Miños Fayard*120
- Un sistema de evaluación con múltiples componentes para sostener el aprendizaje. Cátedra de Matemática,
Facultad de Arquitectura por *V. Albarenque, S. Ballester, P. Birriel, I. Bouvier, S. Donadini, O. Echenique, O. Gil, F. Giménez, G. Ibarburu, A. Nuin, M. Perchman, E. Prieto, M. Reymondo, G. Ríos, L. Rosa, R. Sellanes*122
- Evaluación formativa en cursos de bases biológicas de Odontología. Unidad de Apoyo a la Enseñanza, Facultad de Odontología por *M. Seoane*123
- Experiencias sobre evaluación formativa en Pediatría, Facultad de Medicina, por *M. Sosa*125
- Formatos de evaluación del Proyecto Integral “Protección social, instituciones y práctica profesional”. Licenciatura en Trabajo Social por *E. Ortega, C. Bentura, A. Tommasino, A. Mariatti, C. Saprizza, M. Beltrán*127
- Evaluación formativa en el primer año de Medicina por *M. Esteche, P. Hernández*129
- Experiencias de evaluación formativa en la Facultad de Derecho por *J. López Portillo, M. Casaña, M. Leles da Silva*130
- Evaluación formativa del practicantado de Producción Animal de la Facultad de Veterinaria, Udelar por *C. Morón, R. Kremer, J. L. Repetto, D. Cavestany, C. Borlido, J. Passarini* 132
- La incorporación de la evaluación estudiantil de la docencia en el IENBA por *S. Artigas*134
- Experiencia de evaluación en el Instituto Superior de Educación Física por *M. Sarni, I. Pastorino*136
- Ética y profesionalismo en la enseñanza clínica de la ginecología: experiencia en simulación en estudiantes de grado y posgrado por *G. Curbelo, Cóppola, L. Briozzo*138
- Experiencia en la Facultad de Ciencias Económicas y Administración por *A. López Murissich*140
- La evaluación entre pares. Experiencia del ciclo introductorio, Facultad de Medicina por *Arillaga, Abelenda*142

Presentación

La evaluación en la educación superior siempre ha sido un asunto crítico por su papel certificador de las prácticas profesionales. No obstante, es claro que en las últimas décadas estos procesos se han complejizado notablemente en función de la envergadura de las transformaciones ocurridas en las universidades. Hemos visto en estos años cómo se han introducido cambios significativos en las prácticas de evaluación, diversificándose los tipos de instrumentos empleados, la naturaleza de las tareas solicitadas a los estudiantes y las condiciones de administración de la evaluación. Además, resulta evidente la importancia creciente que adquiere el empleo de los resultados de las evaluaciones como indicadores de la calidad de la enseñanza en los sistemas de evaluación institucional.¹

La proliferación de evaluaciones diagnósticas aplicadas en etapas clave de los procesos de formación universitaria, con finalidades diversas, son un claro reflejo del auge de las políticas de aseguramiento de la calidad.

En las orientaciones de política educativa adoptadas por la Udelar en la Ordenanza de Estudios de Grado se define por primera vez con carácter normativo la función formativa de la evaluación de los aprendizajes, trascendiendo su tradicional función verificadora y certificadora de conocimientos.²

.....

1 Camilloni, A. Basabe, L. y Feeney, S. (2009) *Los formatos de evaluación de los aprendizajes y sus relaciones con las modalidades de estudio de los alumnos universitarios. Perspectivas de investigación y marcos de análisis. Ponencia presentada en el Primer Congreso Internacional de Pedagogía Universitaria. Secretaría de Asuntos Académicos de la Universidad de Buenos Aires. Septiembre de 2009.*

2 *Ordenanza de Estudios de Grado y otros Programas de Formación Terciaria, 2011, art.37.*

De este modo, la Universidad se alinea con las definiciones didácticas más modernas de evaluación, otorgando un papel esencial a la generación de información durante el proceso educativo con el fin de mejorar las actividades de enseñanza y de aprendizaje, esto es, habilitando la posibilidad de que tanto docentes como estudiantes puedan tomar decisiones de mejora de forma oportuna y no sólo una vez culminado el curso. La investigación en educación nos muestra además, que para mejorar sus aprendizajes los estudiantes necesitan retroalimentación frecuente y rápida, a la vez que poder autoevaluarse.

Estos nuevos enfoques resultan prioritarios para un nivel de la educación que tiene como uno de sus propósitos fundamentales la formación de profesionales autónomos y reflexivos.

En este contexto, la Comisión Sectorial de Enseñanza (CSE) ha organizado, junto con la Red de Unidades de Apoyo a la Enseñanza, tres Jornadas sobre la temática en cuestión. En cada oportunidad contó con el aporte de destacados especialistas en la materia, de reconocida trayectoria en universidades públicas de la región y, de docentes de la Udelar que presentaron sus experiencias y reflexiones críticas.

La primera Jornada titulada *Dilemas, conflictos e interrogantes sobre la evaluación de los aprendizajes en la Universidad* se desarrolló en setiembre de 2011. En el marco de la discusión sobre la Ordenanza de grado, la CSE convocó a docentes y estudiantes a debatir a punto de partida de la conferencia dictada por Alicia Camilloni, Profesora Emérita de la Universidad de Buenos Aires (UBA) y de las ponencias de representantes de la Asociación de Docentes de la Universidad de la República (ADUR), Prof. Laura Coitiño y, de la Federación de Estudiantes Universitarios del Uruguay (FEUU), Bach. Carolina Cabrera.

Al año siguiente, octubre de 2012, se desarrolló la segunda Jornada sobre evaluación. En esta oportunidad la temática se centró en: *Las pruebas diagnósticas a debate*. Los trabajos estuvieron a cargo de la arriba mencionada Prof. Alicia Camilloni, el Director de Evaluación y Estadística de la ANEP, Prof. Andrés Peri, la Directora de la Unidad de Enseñanza de la Facultad de Ingeniería de la Udelar, Prof. Marina Míguez, organizados en una Mesa redonda con la moderación del Director de la Licenciatura en Ciencias de la Comunicación de la Udelar, Prof. Gabriel Kaplún. La instancia fue seguida de un nutrido intercambio con el auditorio.



Finalmente, en mayo de 2014, la Jornada *¿De qué evaluación formativa hablamos?* contó con la conferencia de la Prof. Rebeca Anijovich, docente de la UBA, de la Universidad San Andrés (UDESА), Facultad Latinoamericana de Ciencias Sociales (FLACSO) y con comunicaciones de experiencias de docentes de distintos servicios y áreas de conocimiento de la Udelar.

Las conferencias, las ponencias presentadas y los debates generados en las tres Jornadas sobre estas distintas vertientes de la evaluación en la educación superior, transcritos y reunidos en este libro es el material que ofrecemos al lector.

Mercedes Collazo

Primera parte

Dilemas, conflictos e interrogantes sobre la evaluación de los aprendizajes en la Universidad

Este encuentro se realizó el
1.º de setiembre de 2011 de 14 a 18 h
en el aula Magna de la Facultad de Psicología
(Tristán Narvaja 1674 - Montevideo)
Organizado por:
CSE - Unidad Académica
Red UAE

Mesa redonda integrada por:
Apertura
Dr. Luis Calegari
Prorector de Enseñanza
Bach. Carolina Cabrera
Federación de Estudiantes Universitarios del Uruguay (FEUU)
Prof. Laura Coitiño
Asociación de Docentes de la Universidad de la República (ADUR)
Alicia Camilloni
Prof. Emérita de la Universidad de Buenos Aires (UBA)

Apertura

Luis Calegari

Muy someramente vamos a referirnos al sentido de esta instancia, que creemos que es el inicio de una reflexión, de una discusión amplia sobre evaluación de los aprendizajes en la universidad. Todo esto surge a raíz de la discusión de la ordenanza de grado, una herramienta

para empezar a trabajar muy fuertemente en la revisión de los planes de estudio, de las metodologías y de tantas otras cosas que precisamos para renovar la enseñanza de la universidad.

A raíz de la discusión de la ordenanza, que tiene algunas pautas generales de evaluación, pero fundamentalmente cuando se discutía alguna pauta general de calificaciones, notamos las grandes confusiones que todavía imperan sobre todo en los cuerpos docentes sobre estas, sobre la concepción de la evaluación, sobre las formas de evaluación; en definitiva cómo evaluamos. Los estudiantes estudiaron muy profundamente el tema y nos estimularon a pensar en esta primera instancia y de aquí en más empezar a discutir sobre la evaluación, en el entendido de que es imprescindible que lleguemos a acuerdos, que entendamos de qué estamos hablando y podamos definir entre todos conceptos y pautas generales de evaluación, porque de la mano de esto viene la calidad de la renovación de la enseñanza que podamos llevar delante de aquí en más.

Para nosotros es un tema sustancial y para ello hemos invitado a Carolina Cabrera, de la FEUU, que nos va a expresar dónde ven ellos los «nudos» vinculados a la evaluación, y a Laura Coitíño, de ADUR, por el orden docente, después vamos a tener una conferencia y una charla con la profesora Alicia Camilloni, que ha investigado sobre este tema, y finalmente vamos a tener una discusión en plenario para ver cómo seguimos de aquí en adelante.

Carolina Cabrera (FEUU)

Me llamo Carolina Cabrera, soy delegada de la FEUU al Consejo Directivo Central. Me interesa particularmente mencionar por qué surgió la idea de hacer esta charla, esta conferencia, en particular con el prorector de Enseñanza y con Mercedes de la Unidad Académica de la Sectorial. Lo habíamos conversado bastante, porque de cara a la Ordenanza General de Grado la Asamblea Nacional de Claustro resultó bastante engorroso el tratamiento de la propuesta de cambio de escala de calificaciones, y nos sorprendió particularmente —no me voy a referir a ningún orden en particular ni a ningún servicio— cómo en aquella discusión que hacía referencia estrictamente a un cambio en

la escala de calificaciones se terminó armando un nudo tremendo y englobando muchas cuestiones que tenían que ver con la evaluación y con el proceso general de evaluación de cursos y actividades en la universidad con la nota que tenían los estudiantes o que deberían recibir los estudiantes al fin de un curso.

Obviamente el tema de la escala no se pudo seguir tratando como tal porque había un montón de cuestiones enredadas con eso y ameritaba seguir discutiendo conceptos previos o que iban de la mano con la propuesta de nueva escala pero que no tenían que ver estrictamente con el sistema numérico de la escala. A raíz de esa discusión y después a raíz de una reflexión de la FEUU y de recabar distintas opiniones y distintos problemas de todos los estudiantes de la universidad, de muy diversos servicios de la Universidad de la República, se evidenciaba que había en la carrera de esos estudiantes, cada vez que se sometían a un proceso de evaluación, problemas que referían a un montón de cosas.

En primer lugar y en particular a nivel central de la universidad, confundir la escala de calificaciones con la evaluación de los cursos, mostraba una discontinuidad, un divorcio entre la escala de calificaciones y la evaluación de los cursos, al pensar que la nota final que tiene un estudiante refleja el proceso que hizo, cuando en general no es así. Por distintas razones, en distintos servicios muchas veces se nombra el problema de la calificación —aunque no es aplicable en todos los servicios de la universidad— el poner como resultado final del proceso de enseñanza-aprendizaje de los distintos estudiantes en los cursos o en los años enteros de su carrera la nota que reciben el día que van al examen, cuando en realidad muchas veces o no fueron sometidos a un proceso de evaluación previa o fueron sometidos a dos o tres procesos de evaluación durante un curso, pero lo único que cuenta es cómo le fue ese día o cuán motivado estaba ese día para hacer esa prueba.

Fijar las pautas de la evaluación de acuerdo a lo que deberían saber los estudiantes o cuánto deberían haber fijado para ese día al acercarse a las instancias de evaluación, sean pruebas parciales o exámenes finales de los cursos, sin saber realmente lo que eso significa, con lo que se van a enfrentar, suele ser un problema bastante evitable, en primer lugar, pero bastante frecuente. Lamentablemente, los estudian-

tes se enfrentan a evaluaciones finales de los cursos sin saber cuáles van a ser las pautas, sin saber si estudiar para una prueba de múltiple opción oral o escrita. Aunque parezca disparatado, muchas veces los estudiantes llegan a las instancias finales de evaluación sin saber absolutamente nada: cuántas preguntas van a tener que responder, qué tipo de preguntas van a ser, si se va a evaluar todo el contenido del curso o no. Son cuestiones básicas que deberían ser evitables pero que lamentablemente ocurren a menudo.

Quiero resaltar la buena oportunidad que significa estar discutiendo esto hoy aquí. Y el desafío que implica en el contexto actual de discusión de una ordenanza de grado aprobada hace dos días, de un contexto en el que se habla tanto sobre todo lo que tiene que ver con desvinculación, y justamente con la desvinculación sobre todo al ingreso de las carreras, que también es una discusión reciente y en boga y que creo que todos los servicios están llevando adelante en este momento, sobre todo vinculada con la necesidad de formación docente y de evaluación docente que surge sistemáticamente como opinión de todos los servicios de la Universidad de la República. En este caso creo que esa discusión se está dando de una manera mucho más sincera, porque no se está hablando ni de formación docente ni de evaluación docente desde una perspectiva punitiva, sino como una manera de mejorar la calidad del proceso de enseñanza, que creo que es lo que debe primar en la discusión, tratando de construir en colectivo, no para penar a nadie, sino como una manera de construir y definir el lugar garantizando una calidad de los estudios universitarios.

Laura Coitiño (ADUR)

Algunos de los elementos que mencionó Carolina estaban básicamente en la idea de cómo arrancar esta breve presentación para lo que debería ser un debate. Necesitamos reflexionar en colectivo y sobre todo necesitamos llegar a conclusiones que nos permitan transformar la práctica de los docentes, en particular en lo que tiene que ver con la búsqueda de calidad. Muchos de nosotros hemos visto referencias relativamente recientes a un estudio que se publicó, que no tiene que ver con la universidad pública en particular, sino con todo el sector uni-

versitario, que señala que hoy el hecho de tener un título universitario no se convierte en un trabajo transformador de realidades, ya que para mucha gente que está accediendo a los estudios universitarios eso no se traduce en un acceso a cierta calidad de vida, a ciertas condiciones de trabajo.

Ese es un elemento que al orden docente le preocupa particularmente, porque va de la mano de esa gran palabra que todos buscamos, que damos por obvia en la formación universitaria, que es la búsqueda de la excelencia, de la calidad. Tenemos que empezar por ahí, ese es el primer punto, la primera consigna a partir de la cual empezamos trabajando. Pero en algún momento se hace necesaria la acreditación de esa calidad, y allí entra el componente evaluación en una de las acepciones más clásicas que tenemos. Tiene que ver con lo que decía Carolina en el sentido de asignar algún tipo de calificación, algún tipo de juicio de valor que reconoce algo. En el orden docente, esa es probablemente la visión más inmediata en la práctica de evaluación del desempeño y el rendimiento académico de los estudiantes.

El otro componente que a veces está mucho más oculto es que la mayoría de los docentes llegan a la actividad a nivel universitario sin tener una preparación en los temas de evaluación y se refieren simplemente a su práctica, probablemente de los tiempos de estudiante, y a su propia experiencia. Llega con ese bagaje, de esa manera reproduce lo que vio que funcionó para él y lo traslada en un escenario probablemente desfasado en el tiempo, desfasado en las condiciones, esa es la forma en que lo afecta.

Entonces como primer elemento para el orden docente está la preocupación por tratar de evidenciar, poner en la superficie estos temas y sobre todo explicitar. Muchas veces el docente directamente va y hace la tarea de evaluación, y allí lo que tenemos es todo riesgo. Carolina mencionaba esa práctica que tenemos de trabajar con formas de evaluación sumaria. Eso está mucho más ligado a esa necesidad de acreditar conocimientos, pero no tiene el otro componente, que es el que desde el punto de vista de la búsqueda de la calidad sería el más importante para la marcha del trabajo que tenemos día a día como docentes, que es tomar la evaluación como una instancia más de aprendizaje en primer lugar, y además como un espacio para

incidir sobre los procesos que están teniendo los estudiantes para favorecer ese acercamiento a la calidad máxima, pero que son esos trayectos que hacen dentro de la universidad para llegar a determinadas acreditaciones.

Podemos plantearnos —volviendo al título de esta convocatoria— cuáles son todos los dilemas, y los dilemas son muchos, algunos de ellos probablemente falsos dilemas, cuando nos tenemos que plantear si vamos a evaluar para acreditar o vamos a evaluar para formar y para apoyar esos procesos. Las dos cosas están presentes. Tampoco podemos poner el énfasis absoluto en ese trabajo acompañado del estudiante y olvidar que sí hay un rol de la institución universitaria de acreditar, de dar un reconocimiento a las distintas etapas que van recorriendo los estudiantes.

Todas estas son preocupaciones que probablemente presenten distintos matices, porque el orden docente a lo largo y ancho de la universidad tiene muchas variedades, muchas situaciones muy diferentes, que tienen que ver con las especificidades de las carreras, de los distintos servicios. Eso hace que probablemente necesitemos poner una mirada muy específica en cada campo, o sea que si bien hay que llegar a consensos generales, también hay que respetar esas diversidades. Esa es la mejor manera de atacar los problemas que podamos tener, que son los que plantea el segundo ítem del título, los conflictos, todas esas tensiones. No olvidemos que en la medida en que se trata de una evaluación también hay juegos de poder detrás de todo esto y ese es otro problema del que se habla poco pero que tiene que estar sobre la mesa siempre. Esto nos lleva a la necesidad de transparentar las reglas de juego al principio, claras para todos, pues es la mejor garantía para el docente, para ejercer su trabajo de formar, de hacer el acompañamiento y de evaluar, pero también es la mejor garantía para el estudiante que llega a ese curso saber cómo va a tener que enfrentarlo y cuáles son las expectativas que se le plantean.

Las expectativas conforman otra dimensión. Muchas veces los docentes se encuentran a mitad de agua entre dos extremos: qué es lo que piensa el estudiante que tiene que ser una evaluación y cuáles son las expectativas de la propia institución, desde el lado académico como organización de lo que debería ser un proceso. Necesitamos puntos de

referencia. La reglamentación es: sin restringir lo que es la diversidad ni poner un corsé a la libertad del docente de ir cambiando rumbos mientras está hecho el trabajo, tener puntos de referencia comunes, porque somos todos parte de la misma universidad y todos atendemos a un concepto y a las distintas expresiones de lo que es la calidad en una enseñanza universitaria.

Por ahí hay bastantes puntos. Espero que después de la intervención de Alicia podamos arrancar con el debate, ese sí es un pendiente que se siente, porque en general logramos hacer las presentaciones, logramos presentar la información, pero la instancia de debate queda pobre y no avanzamos en ese terreno.

Nudos de debate sobre la evaluación de los aprendizajes en la universidad, necesidades de actualización, profundización e investigación educativa

Alicia Camilloni

Ustedes han planteado, estudiantes y docentes, muchos temas que han sido tratados, que son tratados en el campo de la evaluación. Las cuestiones respecto de las teorías y las prácticas de la evaluación ya tienen más de cien años de investigación. No es un tema que no haya sido cuidadosamente reflexionado, pensado. Se ha producido muchísimo material tanto en forma de proyectos de evaluación, de diseños de evaluación, como de investigaciones en el campo de la evaluación, y uno de los campos que han sido muy trabajados es el de la evaluación en la enseñanza universitaria. Cuando tenía que pensar en qué temas iba a tratar me costó seleccionar, porque todas las cuestiones que tienen que ver con la evaluación son cuestiones debatidas, no es que haya algunos pocos debates localizados.

En primer lugar, las prácticas de la evaluación surgen como la parte visible de un iceberg teórico terriblemente difícil de recorrer, pero además, se caracterizan en todos los niveles de la enseñanza, de la educación formal, porque de todas las actividades que realizan los docentes son las que están más reglamentadas, más reguladas. Nosotros podemos decidir muchísimas cosas: qué vamos a enseñar, en qué

orden lo vamos a enseñar, tomamos muchísimas decisiones respecto de estrategias de enseñanza, pero lo que se refiere a la evaluación está pautado. Tenemos tablas de calificaciones, tenemos regímenes de promoción establecidos y en algunos casos se avanza aún más en la cantidad de instancias de evaluación que deben ser cumplidas, y cuando ya se decide cuál es el régimen de promoción, qué se va a hacer con la nota, el docente se tiene que ajustar a algo que ha sido establecido en una reglamentación. Eso pasa siempre con los docentes, pasa en las escuelas primarias, pasa en las escuelas secundarias y pasa también en la universidad. Por eso me resulta difícil comenzar simplemente con el tratamiento de cuestiones que son claves como la escala de calificaciones y los regímenes de promociones.

Sin embargo cuando se empezó a trabajar en evaluación, por 1908, 1910, el tema que surgió inmediatamente fue qué pasaba con las calificaciones que recibían los estudiantes. Los estudiantes habían venido haciendo ya algunos reclamos, por ejemplo, que los exámenes fueran escritos y no orales. Si eran orales estaban a merced del arbitrio de los profesores —y hubo manifestaciones en la calle y ese tipo de cosas, en ciudades europeas por ejemplo— y pensaban que si eran escritos el docente se iba a avenir a revisar e incluso a justificar la calificación. Cosa que no sucedió, hasta hoy no siempre sucede, a veces sí, y ha sido una buena práctica.

En los comienzos esas investigaciones demostraron que era necesario ocuparse de la evaluación y ocuparse científicamente de la evaluación de los aprendizajes, porque cuando investigaron qué ocurría con las calificaciones descubrieron que el altísimo grado de subjetividad que tenía la calificación que asignaba el docente hacía que fuera incomprensible, incomparable, inesperada totalmente, injustificada la calificación que los estudiantes recibían. No les voy a contar todas las investigaciones que se hicieron en este terreno, comenzaron haciéndose en Estados Unidos, después se asociaron los francófonos, los belgas y los franceses, que usaban muchísimo los exámenes como instrumento para decidir cuestiones que tenían que ver con la vida de las personas, entonces se empezaron a comparar calificaciones asignadas por distintos profesores.

Por ejemplo, la respuesta de un alumno a una prueba de geometría enviada a 147 profesores para que la calificaran entre 0 y 100 puntos recibió entre 29 y 92. Con lo cual muchos de esos profesores la habían aprobado y muchos, en cambio, la habían desaprobado. La misma prueba, la misma respuesta. Se multiplicaron las investigaciones en este sentido. Cuando los franceses se asociaron a este tipo de investigación se produjo un informe clásico en 1936, que demostraba que no solamente en las humanidades las diferencias de criterio de los profesores daban lugar a distintas calificaciones, sino que eso también ocurría en las matemáticas y en todos los niveles de la enseñanza, en la escuela primaria también, la misma prueba de aritmética en la escuela primaria recibía una variedad impresionante de calificaciones.

Esas investigaciones se siguieron haciendo hasta hoy, y lo muy interesante —quiero demostrar que el problema no es sencillo de resolver— es que en una variedad de situaciones, moviendo distintas variables sobre esta cuestión, los resultados que se obtenían en 1910 son iguales a los que se obtienen hoy. Es decir que cuando se sigue entregando a los profesores una pila de pruebas para calificar y los profesores las califican, el orden en que las califican influye en la calificación que ponen y cuando se invierte el orden esas calificaciones cambian. Cuando se coloca una buena prueba en medio de la pila, las pruebas que siguen bajan la calificación. Cuando se coloca una prueba mala en el medio, las que siguen suben la calificación. Esto hasta hoy, por ejemplo una investigación que tampoco es tan antigua muestra que cuando el profesor corrige las pruebas pone una calificación, y si pasado algún tiempo a esas pruebas se les asigna una foto —puesta al azar totalmente, no es el factor de la prueba—, las calificaciones cambian.

En los medios anglosajones la solución a este problema específico hizo pensar que si se utilizaban pruebas objetivas, pruebas que tenían una única respuesta correcta, la subjetividad del profesor, la discrecionalidad, no iba a actuar. Inventaron las *pruebas objetivas*, por ejemplo la opción múltiple, verdadero/falso y una variedad de pruebas objetivas, pensando que eran la solución al problema. Para los belgas, franceses y suizos francófonos, que también se asociaron a estas investigaciones, esa no era la solución. Ellos siguieron pensando que había que encontrar un modo que fuera justo, equitativo, previsible de calificación de

pruebas, aunque estas pruebas no fueran las llamadas pruebas objetivas. Es decir que el problema sigue existiendo, solo que las soluciones aparecen de manera diversa, según la cultura institucional, según lo que se piensa que es el aprendizaje y lo que los alumnos tienen que aprender.

Puesto esto acá, lo que quiero decir es que las discusiones sobre la calificación son una cuestión que surge constantemente. Encontraron por ejemplo que algunos profesores usaban los grados superiores de la escala, otros usaban los inferiores, unos usaban los pares, otros usaban los impares, y que si vemos qué hacíamos nosotros con las calificaciones lo más probable es que encontremos ese tipo de curiosidad. No fueron fáciles de resolver, hasta hoy no ha sido fácil, a pesar de que se han ido mejorando mucho las ideas, las propuestas, las concepciones acerca de qué es la evaluación de los aprendizajes.

Entonces, sabiendo que este no es un tema fácil de resolver, voy a tomar algunas cuestiones más de base respecto de la evaluación, porque si no, se va a convertir simplemente en una cuestión de discusión de mediciones. No es una discusión menor. Cuando vamos a evaluar, cuando vamos a emitir un juicio de valor, lo hacemos en función de una escala de calificaciones, que es un tipo de expresión de medición. ¿Qué es lo que medimos?, ¿cómo se mide?, ¿se puede medir? Todas estas cuestiones dependen mucho de lo que uno piense, si lo que el alumno sabe puede ser medible, puede ser medido con una especie de regla numérica, con una escala de calificaciones compuesta por números, una escala discreta, en donde hay saltos, y que esos saltos y ese número al cual uno llega representan algo.

Voy a tratar de hacer un paneo de algunos de los temas que hoy están en discusión en el campo de la evaluación y para los cuales la enseñanza universitaria necesita recurrir a una apreciación muy cuidadosa, porque probablemente el principal problema que uno encuentra es la separación entre la evaluación y la enseñanza. Una muy buena especialista, Linda Jefferson, dice que nunca habría que hablar de evaluación sin hablar al mismo tiempo de la enseñanza. No se puede dar un curso de evaluación sin dar al mismo tiempo un curso de enseñanza, porque lo que interesa es la vinculación entre las dos cosas, que es más o menos a lo que habría llegado como conclusión. Por tanto voy

a tomar algunos de estos temas para presentar los grandes nudos en los cuales uno tendría que trabajar y en los cuales mucha gente está trabajando.

En primer lugar, una de las cuestiones importantes en este tiempo en la enseñanza universitaria es poder encontrar la relación cierta entre el enfoque pedagógico al que responde el proyecto de formación que da la universidad, el currículo, y los modos, las modalidades con los cuales se evalúa lo aprendido por los alumnos.

Por la década de los 60, que era la época de las planificaciones, cuando las soluciones venían de las planificaciones, empezó a difundirse mucho en el medio de la educación la expresión «erosión de los objetivos». Porque si uno miraba lo que decían los planes de estudios, los programas de las materias, y después revisaba lo que se evaluaba, no había ninguna relación entre esos enunciados ampulosos que aparecían en los planes y lo que efectivamente después se aprendía. No sé qué era lo que se aprendía, sí se podía ver lo que se evaluaba, que era una minimización, una caricaturización, se hablaba de la erosión, justamente. Quedaba muy poco de aquello que se pretendía que los alumnos aprendieran.

Necesitamos garantizar que cualquier solución que uno encuentre en el campo de la evaluación sea realmente representativa del proyecto pedagógico de formación que tiene la universidad. Otra cosa no es aceptable, no tiene sentido, y no lo tiene por algunas razones que voy a ir mencionando.

En buena medida, hoy para tratar de solucionar el problema se está acudiendo a un tipo de diseño del currículo, por competencias. Es una discusión que hay que dar, sobre todo cuando uno discute cuál es la definición de competencia que va a utilizar. ¿Va a ser una definición muy precisa, muy específica, de saberes absolutamente específicos para situaciones determinadas? ¿Vamos a hacer una formación mucho más general, más abierta, de un profesional que se prepara para un futuro incierto? Ese es el futuro en el cual hoy tendríamos que colocarnos.

Mejorar la evaluación no es algo que debiera hacerse a expensas del empobrecimiento de la formación. Esta es la cuestión, hay cosas que yo puedo evaluar fácilmente, pero ¿voy a limitarme a enseñar aquello que se puede evaluar fácilmente? Esta ha sido una solución, y es una

solución de peligro. A mi juicio, es una solución que nos pone frente a un abismo. Porque lo que nosotros tenemos que hacer es formar a la gente en las cosas que pensamos que tiene que saber y en la variedad de saberes a los cuales tiene que llegar, y no simplemente en aquello que nos resulta más fácil de evaluar, medir.

Entonces, ¿cómo se evalúa? Necesitamos un currículo flexible y que plantee muy seriamente las cuestiones de dilema entre la integración y la fragmentación, porque por el lado de la facilidad en las organizaciones académicas que requieren los programas de formación, la universidad, como en general los sistemas educativos, ha recurrido a la fragmentación: distintas personas son responsables de enseñar diferentes cosas y evaluar pedacitos diferentes. Pero necesitamos una formación integral, los problemas profesionales no se resuelven monodisciplinariamente, no vamos a formar un profesional enseñándole pedacitos sueltos. En cada una de las materias le damos un pedacito, pero eso no es formar y necesitamos pensarlo en términos de integración. Claro que es más fácil evaluar en cada una de las materias, en cada una de las materias cada una de las unidades, en cada una de las unidades cada uno de los temas. Entonces el alumno puede saber un tema, pero otro tema no o puede saber cinco temas pero sin ninguna vinculación entre ellos. Cuando encaramos la evaluación de esta manera, le estamos fragmentando el saber al alumno, y eso no es bueno.

Necesitamos pensar en términos de enseñar de manera integrada y de evaluar de manera integrada, pero por razones organizativas de distribución de los tiempos en la formación tendemos a enseñar fragmentariamente y a evaluar fragmentariamente, por eso tenemos un currículo con unidades separadas, enseñanza con unidades separadas y evaluación por pedacitos de trabajos. Es necesario encarar esto de otra manera, pero hay cuestiones organizativas en la organización académica y no solamente en el diseño curricular. Un diseño curricular reclama una forma de organización académica alineada con la organización académica, el diseño curricular, la enseñanza y la evaluación.

En este sentido es que abogo por una fuerte integración de este tipo de cosas, pero al mismo tiempo, si me coloco en el nivel del diseño curricular y pienso en términos de evaluación, ocurre que solemos partir del perfil de un graduado. Este es un recurso de procedimiento formal

para que se aprueben los graduados. ¿Qué tipo de graduado queremos en función de los objetivos? Si reuniéramos todas las pruebas de evaluación que el estudiante pasó a lo largo de la carrera en la universidad, ¿eso sería representativo del perfil del graduado y del proyecto pedagógico que tiene la universidad? Esa es la pregunta que nos tenemos que hacer. Cuando el profesor enseña su disciplina, ¿parte de su disciplina, piensa en términos de la contribución que tiene esa disciplina a la formación del graduado final que queremos? Cuando evalúa, ¿piensa en términos de esa contribución al perfil del graduado? ¿O lo piensa exclusivamente como marco de referencia a su disciplina, su responsabilidad básica, que es que el alumno aprenda esa disciplina, y no en términos de formación general?

Sin embargo, si bien esto hay que hacerlo, hay que pensarlo cuidadosamente y hay que pensarlo como proyecto integral de quienes están desarrollando una carrera, porque no tiene solución cátedra por cátedra o asignatura por asignatura, sino que es necesario como proyecto integral, también hay que recordar que así como hay objetivos finales —el perfil del graduado, incluso las incumbencias que por ley se establecen, porque la universidad está formando profesionales, aun pensando en eso que sería lo final—, también hay objetivos intermedios, entonces no es sencillo definir qué es lo intermedio, qué es un paso para, y cuáles son los finales. Así que no solamente podríamos definirlo en términos de «ha logrado tal competencia» final, sino que necesitamos pensar cuáles son los pasos intermedios que nos llevan a eso. Este tipo de cuestiones se relacionan con diseño curricular y enseñanza y organización académica.

Las relaciones entre el enfoque que uno adopta en la enseñanza y las formas en que evalúa no siempre son coherentes. Viendo lo que ocurre en general en distintas universidades, en distintas carreras, me encuentro con profesores que adoptan un modo de enseñanza pero que a la hora de evaluar, el formato de la evaluación, la modalidad de la evaluación, no responde a la modalidad con la que habían enseñado. Esto no es deseable, hay muchas maneras de evaluar y uno tendría que encontrar coherencia entre ellas. Cuando no hay coherencia, el estudiante no sabe qué se le va a preguntar, no sabe cómo se le va a preguntar, por lo tanto tiene un problema en la preparación, en el estudio. Y

cuando el estudiante no sabe cómo estudiar todo puede ocurrir, porque cada uno de los formatos de evaluación evalúa cosas diferentes, saberes diferentes, modos diferentes de expresar el saber.

Esto ha sido claramente comprobado, hay estadísticas muy interesantes que muestran por ejemplo que hay estudiantes que aprueban una prueba objetiva, pero si se les toma sobre ese mismo tema lo que se llama una prueba de ensayo en la que tienen que desarrollar el tema, tienen libertad para hacerlo, bajan la calificación, en los términos de ese profesor es inferior. Es decir que si el profesor enseña diciendo «yo quiero que los alumnos pregunten, construyan, descubran», etcétera, y después les toma una prueba objetiva en la que espera que den ciertas respuestas, lo que se llama problemas *bien definidos*, problemas que tienen una única respuesta correcta, esos estudiantes van a tener dificultades para trasladar esa respuesta correcta, porque se les había mostrado un panorama diferente. La inversa también existe: el profesor explica cuáles son las pautas con las cuales deben ser respondidas las cosas y qué camino hay que seguir para encontrar la respuesta correcta, y después plantea cuestiones abiertas. Termina generando una terrible ansiedad en los estudiantes, que no saben cómo responder ante ese tipo de situaciones para las cuales no habían sido preparados. Entonces una de las normas siempre es: no evalúes de un modo que sea inesperado para los estudiantes, el estudiante tiene que saber cómo va a ser evaluado.

Acá aparece una cuestión fundamental, que la investigación de las universidades con los estudiantes universitarios ha mostrado ya con muchísima claridad: que los aprendizajes de los estudiantes dependen del formato de evaluación que saben que va a ser utilizado mucho más que de la enseñanza que han recibido del profesor. Lo que finalmente el estudiante aprende depende de cómo percibe que va a ser evaluado. Si tengo una prueba de opción múltiple, podría perfectamente construirla de manera que el alumno tenga que hacer un razonamiento bien complejo, pero hay investigación que demuestra que los estudiantes cuando saben que se les va a tomar una prueba de opción múltiple la perciben como que se les van a pedir datos: fechas, nombres, fórmulas, es decir, siempre fragmentos de información, y no piensan que a través de ese tipo de formato se les puede pedir que desarrollen

un razonamiento muy complejo. Entonces se preparan de diferente manera. Cuando ellos saben que se les van a tomar problemas o pruebas de ensayo, problemas de esos que se llaman *mal definidos* porque tienen distintas respuestas correctas, se preparan de una manera, y cuando saben que les van a tomar pruebas de opción múltiple o de ese tipo se preparan de otra, estudian de manera diferente. Si a esos estudiantes que estudiaron así uno les cambiara sobre la marcha el tipo de prueba, los condena al fracaso.

Repito, hay mucha investigación que muestra que el estudiante adecua su estudio a la modalidad con la que sabe que va a ser evaluado. Así que la elección de la modalidad y el conocimiento que tenga el estudiante de cómo va a funcionar esa evaluación son absolutamente fundamentales para el éxito del estudiante. Porque si el estudiante lo ve erróneamente, si lo percibe mal, se va a preparar mal y no va a aprobar.

Pero detrás de eso hay muchas otras cosas, no se trata solo del éxito del estudiante, sino de la coherencia de la formación que queremos que tenga. Lo que importa es que si enseñamos de una manera que no es una manera que recurre simplemente a la memoria mecánica, a la memoria de informaciones fragmentales, sino que nos interesa su razonamiento, la capacidad de producción de conocimiento que pueda tener el alumno, entonces evaluaremos de manera coherente con ese objetivo. No es que cualquier modalidad de evaluación sirva; cada modalidad de evaluación promueve un cierto tipo de actividad cognitiva y por lo tanto evalúa un cierto tipo de actividad cognitiva, no cualquiera, sino que cada una se ajusta a cierto tipo de pautas. Detrás de eso hay otro problema que es cómo percibe el estudiante ese formato de evaluación, que puede no ser igual a la percepción del profesor. Por lo tanto, si el profesor no le aclara qué es lo que espera, si el estudiante no sabe cómo va a ser evaluado, no sabe qué parte del contenido le va a ser tomada, no sabe cómo enfrentarlo.

Hoy definimos la actividad de estudio como el conjunto de actividades que realiza un estudiante con la intención de aprender. El estudiante tiene la intención de aprender cuando se pone a estudiar, pero cuánto tiempo, cuánto esfuerzo y qué tipo de trabajo tiene que realizar, eso depende del formato de la evaluación. Siempre pensamos que

si enseñamos el estudiante aprende, pero no, el estudiante aprende lo que sabe que va a ser evaluado.

Hoy también se advierte que el resultado de esos aprendizajes varía según los estudiantes. Ya en la década de 1980 resultó de varias investigaciones que ante una misma situación de aprendizaje los estudiantes respondían de diferentes maneras y que los resultados de las evaluaciones eran distintos. La escuela sueca generó una idea binaria de lo que podía ser el aprendizaje diciendo que había *aprendizajes superficiales* y *aprendizajes profundos*, y que el hecho de que se diera uno u otro dependía de lo que el estudiante esperaba hacer en esos exámenes. Cuando un estudiante tenía decidido que eso le importaba, que era importante, quería tener un aprendizaje duradero, dedicaba mucho tiempo, mucha atención, construía muchas relaciones alrededor de los temas en los que estaba trabajando y lograba lo que se llamaba un aprendizaje profundo. Cuando el estudiante se proponía pasar el examen («zafar») el aprendizaje era superficial. Entonces apareció que los buenos estudiantes hacían aprendizajes profundos y los estudiantes regulares hacían aprendizajes superficiales, que desaparecían rápidamente.

Cuando la investigación fue avanzando y se fueron mirando con más cuidado las cosas, aparecieron algunas investigaciones muy interesantes en Inglaterra en las que tomaron buenos estudiantes en la universidad, estudiantes con muy buenos resultados en el aprendizaje, y les preguntaron cómo estudian. Ahí se descubrió que el buen estudiante no realiza solamente aprendizajes profundos, sino que también se ajusta a cómo va a ser evaluado y hace una combinación de aprendizajes profundos con aprendizajes superficiales. Lo que les parece importante lo estudian en serio, le dedican mucho tiempo, pero lo que saben que la cátedra pregunta, esas cosas que a veces hay que saber, lo aprenden con un aprendizaje superficial. Eso llevó a crear una tercera forma, un tercer término, concepto, acerca del aprendizaje: el *aprendizaje estratégico*. El aprendizaje estratégico es esa combinación de aprendizajes profundos y superficiales que los alumnos hacen para aprobar un examen y que tiene que ver con el oficio de estudiante, al cual todos seguramente debemos haber accedido y por eso podemos estar acá.

Repito, resolver qué coherencia va a haber entre el enfoque pedagógico de la enseñanza y el enfoque pedagógico de la evaluación no es sencillo, pero merece muchísimo cuidado, muchísima reflexión, porque es decisivo.

Otro tema al que es imposible escapar es lo que se denomina la *evaluación auténtica*. En general siempre hemos pensado que lo que se enseña en las instituciones educativas es el saber, podemos decir el saber escolástico, y que con el saber que aprende en la universidad el profesional que egresa va a poder resolver los problemas reales con que se encuentre. Ocurre que no es así, hay investigación que muestra que no es así, porque no todos los graduados universitarios, aun con buen desempeño académico, están realmente capacitados para encarar los problemas de la realidad, porque los problemas reales son engorrosos. Yo hace un rato dije que hay problemas bien definidos y problemas mal definidos; los problemas son bien definidos cuando uno recibe la información necesaria para resolverlos y tienen una respuesta; hay que encontrar la respuesta, ese es el trabajo, y la respuesta muchas veces tiene que coincidir con la respuesta que el profesor pensó que era la correcta, en cuyo caso uno tiene una buena evaluación. Pero los problemas que aparecen en la realidad no son así, son borrosos, no nos vienen con toda la información y hasta nos vienen con exceso de información. Es decir tenemos que poder seleccionar la información que es útil, que es relevante, que es importante, separarla de la otra, buscar la que no tenemos y encontrar algunas de las vías de solución.

El profesional universitario hoy tiene que resolver problemas no rutinarios, problemas nuevos. Como ya decía Margaret Mead en la década de 1950, resolver problemas que nadie conoce todavía, o como dice el presidente del Centro Argentino de Ingenieros, «tenemos que formar a nuestros ingenieros para trabajar con materiales que no existen todavía». Esos son los graduados que tenemos que formar, y no los podemos formar con problemas artificiales siempre, temas dados de manera abstracta, por ejemplo, o temas referidos a un caso concreto, porque a veces para apurar esto de la abstracción, esta generalidad, uno da un ejemplo, un caso y le parece que con eso el alumno ya comprendió la teoría. Resulta que no es así, para poder resolver esos

problemas reales es necesario que el alumno aprenda a resolver problemas reales.

Cuando yo digo que hoy uno de los grandes temas de la evaluación es lo que se denomina evaluación auténtica es porque *auténtico* quiere decir 'de la vida real'. En estos casos le doy un problema auténtico cuando es un problema de la vida real, lo coloco en una situación auténtica cuando es una situación verdadera de la vida real y no simplemente el problema escolar que uno da normalmente en la enseñanza y que es el que normalmente se utiliza en la evaluación. No quiero decir que no haya que dar este otro tipo de problemas, los que llamaríamos problemas escolásticos. Sí hay que darlos y constituyen una preparación, una etapa de la formación, pero no puede haber formación universitaria si el alumno no tiene que enfrentar problemas reales utilizando la integración del conocimiento, todos los conocimientos que ha logrado hasta ese momento.

La inclusión de formas de educación experiencial en la formación de los alumnos que podríamos llamar extensión, como prácticas sociales, aprendizaje de servicio o ese tipo de cosas, es justamente la ocasión en la que el alumno tiene que enfrentar una realidad, que es una realidad inasible que hay que conceptualizar, no viene ya conceptualizada por el profesor. La evaluación hoy tiene que incluir este tipo de problemas a lo largo de toda la carrera, eso no puede quedar depositado en un internado, en una residencia, en una pasantía profesional, en un período de aprendizaje de servicio, sino que tiene que movilizar toda la formación. No es fácil, pero este es uno de los temas que hoy están conmoviendo los currículos universitarios.

La relación teoría-práctica no se resuelve simplemente enseñando la teoría y después cómo se aplica en la práctica, sino que la práctica es la ocasión para poner a prueba la teoría y luego revisarla. No me olvido, a pesar de todo lo que vengo diciendo, de que lo que la universidad enseña es teoría, porque la teoría es lo generalizable.

Cada una de las modalidades de evaluación permite evaluar algún tipo de saber, y en este sentido cuando vamos a pensar cómo resolvemos nuestros problemas de evaluación nos encontramos con que hay distintos tipos de saberes que los estudiantes tienen que lograr y que de hecho logran, por lo tanto no pueden ser evaluados de una

única manera. Es decir que si utilizamos un solo tipo de instrumento de evaluación repetido, vamos a estar evaluando solamente una forma de producción del conocimiento. Si los estudiantes pueden salvar con un aprendizaje superficial, empezamos a pensar que a lo mejor el formato de evaluación que estamos utilizando no es el más adecuado. Los formatos que nos permiten evaluar el aprendizaje profundo van a inducir a los alumnos a lograr aprendizajes profundos. En ese sentido el conocimiento que le pedimos al alumno implica comprensión.

Voy a tocar un tema que no es sencillo. Estoy mostrando que son muchos los nudos que uno resuelve cuando decide enseñar de tal modo y evaluar de tal manera. De alguna forma está dando respuesta a todas estas cuestiones. Cuando decimos «vamos a evaluar la comprensión que tiene el alumno», tenemos que recordar que los niveles de comprensión son variados. Necesitamos que el alumno comprenda, es decir que entienda cuál es el significado y que no se limite a repetir una información de la que no entiende nada. El problema es que los niveles de comprensión se diferencian, hay lo que se llaman *fenómenos de semicomprensión*. No es que uno o comprende o no comprende, hay cosas que aprendimos alguna vez y que en el curso de la vida fuimos ido aprendiendo cada vez mejor, comprendimos más, cada vez le encontramos más significado a esto que aprendimos. Por ejemplo las letras de nuestros himnos nacionales, seguramente cuando los aprendimos, cuando éramos chiquitos, no entendíamos nada y con el curso de los años hemos ido comprendiendo mejor.

Cuando un estudiante estudia y dice «yo ya sé», ¿en función de qué dice «ya sé»? Porque uno puede saber muchísimo de algo, puede haber entendido de qué se trata, de qué estará hablando, y pasa por encima de conceptos que cree que entendió, porque textualmente le parece que entiende esos significados. Y cuando la precisión del aprendizaje depende de cómo va a ser evaluado, de cómo cree que va a ser evaluado, para eso tiene que estudiar, y cuando nosotros hablamos de evaluación tenemos que hablar del producto de la preparación de estas actividades intencionales a juicio del alumno para aprender. Porque lo que el alumno aprende no depende solo de lo que el profesor hace o enseña, sino de las decisiones que el alumno toma respecto de cómo va a estudiar, cuánto va a estudiar, cuándo se va a detener y va a decir «ya

sé». Yo puedo decir que dentro de quince días voy a tomar un parcial, entonces algunos alumnos se ponen a estudiar ese mismo día y otros dicen «esto en dos días o en unas horas lo aprendo». Para eso el estudiante tiene que conocer sus propias fuerzas.

Cuando el estudiante estudia y dice «ya sé», algo entendió, algo comprendió. Ese estudiante piensa que cuando lee comprende y por lo tanto ese nivel de comprensión es suficiente para aprobar un examen. Todos los estudiantes toman estas decisiones y de esas decisiones depende lo que aprenden. Entonces, en la medida en que no nos interesa la evaluación solo para poner una nota sino para mejorar los aprendizajes, es decisivo para mejorar los aprendizajes que los estudiantes tengan claro qué es lo que se espera que hagan y qué es lo que se espera. Se espera que comprendan. Pero la comprensión no puede estar en el grado uno, dos, tres, cuatro, cinco, salvo que les expliquemos qué es lo que esperamos de ellos es prácticamente imposible que se autoevalúen. Porque la clave de todo esto es que el alumno pueda decir «sí, ahora ya sé». Y vale la pena que lo sepa tanto para poder detener el estudio y decir «ya sé» como para comprender la evaluación que hizo el profesor, para poder discutirla, porque si no sabe qué es lo que sabe y no entiende cuál es el nivel de conocimiento que ha alcanzado, queda marginado totalmente de la evaluación.

Cuando uno toma una prueba de evaluación le puede pedir al alumno que evoque una información, le puede pedir que infiera algo, puede ir un paso más allá, le puede pedir que infiera alguna conclusión o que construya un conocimiento. Pero también le puede pedir que utilice el razonamiento o conocimiento puramente abstracto en general. O puede pedirle que se refiera a casos absolutamente concretos y específicos. También puede pedirle que además de mostrar que posee un conocimiento, demuestre que posee una habilidad y puede ser que demuestre una habilidad para utilizar ese conocimiento, con lo cual hoy se habla de *conocimiento condicional* cuando el alumno no solo posee conocimiento, sino que sabe también cómo, dónde y cuándo utilizarlo. Por lo cual la evaluación debería referirse a este tipo de conocimiento, un conocimiento en el que hay un nivel de comprensión que alcanza para utilizarlo en situaciones adecuadas.

Un ejemplo elemental. Uno les puede enseñar a los alumnos a resolver cierto tipo de problemas. Los alumnos resuelven esos problemas. Se puede suponer que el conocimiento, la información o las habilidades que los alumnos están utilizando para resolver los problemas han sido incorporados por ellos, desarrollados, hechos carne cuando son capaces de plantear nuevos problemas que se pueden resolver utilizando esos conocimientos. No simplemente cuando resuelven los problemas que les dio el profesor, sino cuando son capaces de formular problemas que se resuelven utilizando los conocimientos que han logrado adquirir. Ahí podemos decir que estamos mostrando que hay un conocimiento condicional: saben cuándo, dónde y cómo utilizar ese conocimiento. Obviamente, si les planteo cuestiones de la vida real tendrán que utilizar su conocimiento condicionado.

Hay cuestiones que tienen que ver con la evaluación y que implican que en principio, cuando formamos un grado universitario, es alguien que tiene capacidad de autonomía en la toma de decisiones. Esto es absolutamente fundamental, una de las cuestiones que tenemos que desarrollar en cualquier graduado es la capacidad autónoma de tomar decisiones apropiadas en los campos en los cuales tiene que moverse. Pero ocurre que entre las cosas interesantes que se fueron desarrollando a lo largo del siglo XX está la idea de que hay problemas que uno puede resolver autónomamente, de manera independiente, pero hay problemas que son más difíciles y más complejos, que se tienen que resolver con la ayuda de alguien. Si alguno de los presentes ha escuchado hablar de la *zona de desarrollo próximo* de Vigotsky a lo mejor entiende lo que estoy diciendo. Vigotsky era un psicólogo que estudió —hace ya bastante tiempo, murió en la década de 1930— que uno independientemente, autónomamente puede llegar a resolver cierto tipo de aprendizaje, pero que si es ayudado por otro que no le dice cómo hacer, sino que simplemente lo va acompañando y planteándole algunas preguntas interesantes, puede superar esto y llegar más lejos. Con lo cual parecería que hay dos niveles de saber, de aprendizaje: el aprendizaje al que uno llegó para poder actuar autónomamente de manera independiente, y un aprendizaje mayor en el cual con algún tipo de ayuda uno recorre un camino adicional y va más lento. En general cuando hablamos de evaluación, hablamos de la evaluación del traba-

jo hecho independientemente por el alumno. No se puede copiar, no debería recibir ayuda, tiene que saber hacerlo independientemente, autónomamente. Esto representa una parte del aprendizaje iniciado.

Si ahora lo estoy viendo dentro de un programa de formación, mi enseñanza debiera ayudarlo a atravesar con mi apoyo, con el apoyo del docente, esa otra zona a la cual puede ir accediendo, utilizando una expresión de Jerome Bruner, «el andamio que le pone el docente». Un andamio que luego le va retirar, porque va a llegar a un grado superior de autonomía. Por lo cual acá tengo dos formas de evaluación: una evaluación de lo que puede hacer independientemente y una evaluación de lo que puede hacer cuando va recibiendo algún tipo de ayuda. Esa etapa de la ayuda en la cual uno le va ofreciendo una idea, alguna orientación, algún desafío, alguna pregunta, algún señalamiento, es fundamental en la formación. Es justamente lo que justifica la existencia del docente. ¿Para qué está el docente?, ¿hace falta, si hay tantos materiales a los que se puede acceder? El docente está precisamente para ir apoyando al alumno para resolver algunas cuestiones que le pueden resultar difíciles, algunos obstáculos que no sabe cómo vencer.

Acá aparece esta otra función de la evaluación que no es solamente decir aprobó o no aprobó, sino que es la función de mejoramiento, es la evaluación que hoy se denomina *evaluación formativa*. Es la evaluación que al alumno le da información, que le permite mejorar su aprendizaje. Es la información que recibe cuando está aprendiendo y entonces uno, que no se dedica solamente a ponerle una nota, sino que mira su aprendizaje, dice «en esto está bien, en esto tiene una dificultad, acá hay un error», mira cuál es la razón de los errores, porque un error puede deberse a muchísimas razones diferentes y para solucionarlo hay que saber cuál es la causa. En ese caso uno le va ofreciendo información al alumno para que pueda resolver esa situación y mejorar su aprendizaje.

Hoy sería imposible hablar de evaluación sin hablar de evaluación formativa, es uno de los tipos funcionales de evaluación más importantes. Dos ingleses, William Eiman y Paul Black, hicieron una investigación de metaanálisis, reunieron una cantidad enorme de investigaciones sobre el tema tratando de llegar a alguna conclusión respecto a esto. Se preguntaron: ¿es cierto que el mejor modo de mejorar el

aprendizaje de los alumnos es instalar un sistema de evaluación formativa? Y relevaron investigaciones que se habían hecho con chicos desde el jardín de infantes hasta la finalización de la universidad, miles de investigaciones, y llegaron a la conclusión de que efectivamente el mejor modo de mejorar el aprendizaje de los estudiantes es instalar sistemas de evaluación formativa. Si hacemos una evaluación formativa funciona de tal manera que el alumno en primer lugar debe creer en la evaluación —si no cree en la evaluación la información que recibe del profesor no le va a servir—, en segundo lugar tiene que recibir una información que le sea útil para mejorar su aprendizaje, y en tercer lugar tiene que hacer algo con ella.

Y acá quiero hacer una aclaración. También hay investigaciones que muestran que cuando el alumno recibe su hoja corregida mira la nota y la guarda. Cuando el profesor marca los errores y le devuelve la hoja al alumno, este mira la nota y la guarda. Cuando el profesor se toma más trabajo y le explica por qué está mal, el alumno mira la hoja y la guarda. Cuando el profesor se toma aún más trabajo y le explica qué es lo que tendría que hacer para solucionar esos errores, el alumno mira la nota y la guarda. Quiere decir que el alumno va a guardarla y si se le dan algunas explicaciones escritas también, hay que encontrar otra salida para dar esa información. Tiene que haber una interacción con el alumno, entre el alumno y el docente, en la que realmente haya una explicación, una conversación. Si hoy me preguntan cuál es el formato de evaluación más adecuado, les voy a contestar: la conversación. La prueba se puede usar como ocasión para conversar con el alumno, pero nada sustituye a la conversación.

Hemos convertido la evaluación en una cuestión de formalidades, de ritos, de planillas, de actas, de números. Pero ¿se puede medir con números lo que los alumnos saben? Yo sé que necesitamos números, esto se encuentra también dentro de los problemas de la evaluación, pero una cosa es decir que necesitamos los números y otra es creer en los números, no es lo mismo.

Por eso voy a seguir con algunos temas que tienen que ver con estas cuestiones. Tenemos que evaluar actitudes y no solamente informaciones, porque no hay proyecto de formación de graduados universitarios en el que las actitudes, la responsabilidad, los valores puedan que-

dar afuera. Tenemos que pensar con qué instrumentos de evaluación vamos a evaluar esos aspectos. No podemos quedarnos simplemente en la información y el manejo de información. Necesitamos graduados comprometidos, comprometidos con su profesión, con las personas que van a recibir los productos y servicios de ellos y con la sociedad. Y ese tipo de graduados no pueden haber sido evaluados exclusivamente mediante pruebas que tienen que ver con lo cognitivo. Necesitamos dar una vuelta en el tipo de trabajo que estamos haciendo.

Voy a hablar de un tema un poco más técnico, pero que no puedo dejar de mencionar, que tiene que ver con la construcción de los instrumentos de evaluación. Hay muchos formatos diferentes y algunas cuestiones que son difíciles de resolver.

Voy a empezar por las cuestiones que tienen que ver con la validez. Se dice que una prueba de evaluación es válida cuando permite evaluar eso que uno quiere evaluar con ella. En ese caso se dice que es válida, pero hay varias clases de validez, y una validez importante es la que se denomina *validez de contenido*. Se dice que un instrumento de evaluación tiene validez de contenido cuando representa una muestra significativa del universo de contenido. Quiere decir que tengo el total del contenido del total de los contenidos, quiere decir que contenido es todo lo que enseño. No son solamente informaciones, temas, sino también actitudes, valores, hábitos, destrezas, todo lo que tiene que ver, todo eso es el universo de contenido. Cuando vamos a evaluar es imposible que evaluemos todo, hay un problema de tiempos, no disponemos del tiempo necesario para poder evaluar todo, por lo tanto siempre que vamos a construir pruebas de evaluación vamos a hacer una selección.

La selección se hace de tal manera que uno pueda suponer que el alumno sabe o no sabe eso que no preguntó, eso que no evaluó. Cómo hacer esa selección es uno de los temas más difíciles de resolver. No hay recetas, depende de cada materia, depende de la disciplina, depende de los objetivos que uno tiene en la enseñanza. Cómo seleccionar algunas cosas que colocar dentro de la prueba y otras que dejar afuera, de tal manera que uno pueda suponer que esa nota que le puso al alumno vale para el todo. Todos tenemos la experiencia de «le voy a preguntar esto, aquello, que me parece lo más importante», pero es

lo más importante pero estoy dejando afuera cosas más importantes también. Cuando lo apruebo estoy haciendo una afirmación de que ya sabe, todo lo otro lo sabe. Cuando lo desapruebo, muchas veces me basta con decir «esto es una cosa tan importante que no puedo ignorarla, por lo tanto lo desapruebo», aunque sepa todo el resto. Los alumnos saben muy bien que depende mucho del azar, a veces les toca justo lo que no sabían o justo lo que sabían. Saber hasta qué punto ese saber que demostró en la prueba de evaluación es representativo del total del conocimiento es un tema muy difícil de resolver que requeriría avanzar un poco más y no quedarse en unas muy pocas pruebas de evaluación. Requeriría un trabajo con evaluaciones más frecuentes, que permitan cubrir mayor cantidad de material, mayor cantidad de contenidos. Es muy difícil de resolver, no es algo que uno pueda decir «lo hago y estoy seguro».

Esto y todo lo que vengo diciendo obliga a que todo evaluador sea modesto. Esté dispuesto a poner en duda lo que hace, a autocriticarse, es decir que no se limite simplemente a evaluar al otro, sino que evalúe lo que él está haciendo. Porque son tantas las dificultades que hay que resolver que uno las resuelve por varias razones.

Un tema es la validez de contenido, es terrible como tema, uno casi podría decir que lo más difícil de resolver en la evaluación de los aprendizajes es la cuestión de la validez de contenidos. Otra es la *validez predictiva*, la capacidad que tiene este instrumento de evaluación de predecir dos cosas. Una cosa es cómo será el comportamiento, la durabilidad de los aprendizajes que ha logrado el alumno, ya que hay mucha investigación que demuestra que lo que el alumno sabe un día, un día después parece haberse perdido, y la durabilidad para nosotros es fundamental, porque si el alumno pierde, si ha olvidado lo que sabía al día de la evaluación, ¿qué estamos haciendo? En realidad siempre hay un resto que queda. Se dice que es una curva asintótica: los primeros días hay una pérdida bastante fuerte y después hay algo que se mantiene, que se conserva. Este es un tema, la durabilidad, la perdurabilidad del aprendizaje logrado.

El otro tema es hasta qué punto eso que uno logró sirve como base para realizar nuevos aprendizajes. Porque nuestro objetivo es siempre una formación abierta, cuando enseñamos tenemos que saber que

esto que enseñamos hoy seguramente será revisado, será cambiado, será reconstruido en el futuro, que aparecerán nuevas teorías, nuevos problemas, y que el alumno desde el momento en que se gradúe siempre va a tener que seguir aprendiendo. Por lo tanto, las evaluaciones que hacemos hoy son las de hoy, pero estamos haciendo una apuesta para el futuro también, y es que esta persona va a ser capaz de evaluar lo que sabe y lo que no sabe en su ejercicio profesional para poder seguir encarando su formación, porque sabemos que terminar no es un broche de oro sino que es simplemente una puerta para seguir aprendiendo. Esto es algo que la evaluación tiene que generar, que nuestras formas de evaluación tienen que generar.

Hay otro tipo de validez que se llama *validez manifiesta*. La validez manifiesta implica que las personas que son evaluadas piensan que esa forma de evaluar es correcta, que lo que se está evaluando en esa manera de evaluar realmente indica lo que ellas saben, que son maneras honestas, adecuadas, aptas, para captar sus saberes. Si yo quiero que resuelvan problemas de matemáticas y los hago correr alrededor del salón para ver cuánto tardan en dar la vuelta obviamente no hay relación entre lo que quiero que demuestren saber y mi manera de evaluar lo que saben. Se dice que una evaluación tiene validez manifiesta cuando los alumnos evaluados de esa manera no dicen «hoy se levantó de mal humor», sino que efectivamente esta manera de evaluar y esta manera de calificar son correctas, que la calificación que recibí se justifica, la entiendo, la acepto. Por supuesto, en otros niveles de la educación, en la escuela primaria o en la escuela secundaria, los padres también tienen que creer en esa manera de evaluar, todos los implicados tienen que creer que esa es una manera adecuada de evaluar. Validez manifiesta, validez predictiva, validez de contenido deben ser resueltas en las pruebas de evaluación. No es fácil.

Pero además hay otra característica que preside en buena medida nuestras decisiones sobre la evaluación. Es lo que podríamos denominar *practicidad*. Recuerdo que en algún momento yo decía en algún curso que las pruebas de opción múltiple sirven para evaluar ciertas cosas y otras no, y un profesor me dice: «Yo no tengo otra salida, porque tengo 450 alumnos». Le pregunté cuál era el equipo de su cátedra, cuántos eran, y me dijo: «Un ayudante y yo». Este tipo de pruebas

como la de opción múltiple se han instalado desde hace cien años casi porque son prácticas, porque permiten evaluar a una enorme cantidad de alumnos simultáneamente. Hoy incluso hay sistemas mecánicos con lectoras ópticas que hacen la corrección, entonces se puede corregir rapidísimamente. E incluso hay unos aparatos que no solamente hacen la corrección a través de una lectora óptica, sino que inmediatamente sacan todas las estadísticas. Uno toma la prueba, pasa las hojas y en quince minutos corrige una gran cantidad de alumnos y obtiene los resultados de la evaluación. La practicidad en ese sentido es un criterio del que no siempre nos podemos escapar, porque yo puedo decir que es necesario trabajar con los alumnos, que es necesario hacer una evaluación formativa y que no hay mejor forma de hacerlo que conversando con los alumnos, pero si tengo cursos enormemente grandes tengo que encontrar alguna forma adecuada para poder evaluar, para aproximarme a cuáles son sus saberes e interpretar luego qué pasa mis alumnos. De todas maneras, aun así hay un paso que es mi primer paso de evaluación: obtengo resultados. ¿Qué hago con los resultados?

El segundo paso, que es fundamental, es evitar algo que solemos hacer en todas las escuelas, que es guardar en un cajón los errores que cometen los alumnos; porque con nuestro sistema de calificación le podemos poner por ejemplo 12 si no ha cometido ningún error, 10 si cometió un error y 8 si cometió dos errores. Pero si le pongo 8 aprobó, ¿y qué pasó con sus errores? Los guardó. Esos errores se guardan en un cajón, a veces otro cajón va a tener otros errores y otros, y cuando el alumno termina tiene una cantidad de saberes y una cantidad de cajones con errores.

Por eso hoy la evaluación apunta a que ninguna evaluación se considere efectivamente terminal sin la obligación de que el alumno corrija sus errores, rehaga el trabajo. La evaluación no termina simplemente con una calificación, lo que interesa es lo que se hace después. Yo puedo poner una nota, puedo decir la nota mínima de aprobación es tal, pero ¿cómo distribuyo la calificación respecto de la producción del alumno entre los otros grados de la escala? Como sé que los preocupa el tema de la escala de calificaciones, hasta ahora traté básicamente de demostrar que la evaluación es mucho más que poner una nota. Uno llega a poner una nota después de una cantidad de operaciones impor-

tantes que tienen que ver con cómo enseñó, qué lugar le da a la evaluación, qué formatos elige para evaluar, cómo construye las pruebas de evaluación, qué tipo de devoluciones le tiene que dar al alumno, ¿les tiene que dar una devolución? Pero la devolución no puede ser simplemente una calificación, una nota, un número dentro de una escala.

Hay muchas clases de escalas. Tenemos una escala denominada *escala numérica* porque los grados tienen nombres de números. Esta escala puede ser de 0 a 12, por ejemplo, en toda escala tengo que definir qué cantidad de grados tengo, y cuando defino la cantidad de grados tengo que definir los niveles de calidad a los que corresponden. Si tengo una escala de 0 a 12, tenemos trece niveles de calidad. La nota mínima de aprobación es 3, quiere decir que 0, 1 y 2 son niveles de baja calidad y tenemos diez niveles diferentes de buena calidad. Los niveles de buena calidad pueden definirse de dos maneras: por cantidad de errores, o por profundidad, brillantez de la respuesta, es decir, «esta respuesta está bien, pero aquella otra está mucho más completa», etcétera. Entonces cómo diferencio diez niveles de buena calidad.

Empecé contando que las diferencias de criterio que se hallan en las calificaciones que ponen los profesores en todas las disciplinas en todos los niveles de la enseñanza son notorias y son incurables. Se dice que cuanto mayor sea el número de grados que tiene la escala, menor confiabilidad tiene, menor exactitud, porque cada vez es más difícil definir cada uno de esos grados, y si yo no defino los grados no los puedo ni comunicar al alumno, cuando el alumno recibe un 7 de un profesor y un 9 de otro, ¿sabe por qué uno le puso 7 y el otro le puso 9? Es prácticamente imposible. Para cada materia puede haber definiciones diferentes, para cada profesor puede haber definiciones diferentes y muchas veces no tienen definiciones, con lo cual se convierten en una información absolutamente oscura. Lo que queda claro, como dicen los críticos de la medicina basada en la evidencia, es que calificamos en función de las evidencias que tenemos, y los críticos de la medicina basada en la evidencia dicen que en términos de salud hay solamente dos evidencias posibles: está vivo o está muerto, no hay dudas, o es una cosa o es la otra. Lo mismo pasa con las evaluaciones en educación, o aprobó o desaprobó, salvó o perdió, una de las dos cosas. Cuando empezamos a mirar niveles de calidad, la cosa se hace diferente.

En los estudios investigativos sobre el tema no hay acuerdo en cuándo se aprueba y cuándo no, es más, se muestra que en los puntajes que están alrededor de lo que se llama el *puntaje de corte*, la zona alrededor del puntaje de corte, en este caso el 3, hay muy poca diferencia entre los que están un poco por abajo y los que están un poco por arriba. Es decir que no hay un corte, no hay un designio claro de «este está aprobado» o «este está desaprobado». Salvo que se fije un criterio claro, y justamente la cuestión está en cómo se definen esos criterios, porque el alumno necesita que le den el número. El alumno no comprende si no, y es más, no sabe cómo prepararse, al final, qué es lo que va a aprender.

La clave de una buena evaluación y una buena formación en la universidad —y en cualquier lugar de enseñanza— es la autoevaluación, que el alumno logre autoevaluarse. Porque cuando la evaluación queda como una decisión del otro en la cual uno no cree, ni confía ni entiende, esa evaluación únicamente sirve para poner un número en una planilla, pero no tiene significado para la persona, no tiene significado para el estudiante. El estudiante tiene que comprenderla, este es un primer escalón. El alumno tiene que comprender por qué ha sido calificado de esa manera, y para eso tiene que haber recibido algún tipo de explicación previa.

Pero en segundo lugar, en el otro extremo, ¿puede hoy un profesional, un graduado universitario no ser capaz de evaluar lo que hace, de autoevaluarse? No existe, estaría muy mal formado. Es absolutamente indispensable que sea capaz de autoevaluarse. Esto significa que la autoevaluación no solamente serviría porque le da sentido a la evaluación del profesor, sino que es un contenido que hay que enseñar. Es decir que yo no podría dejar salir a nadie de la universidad con un diploma si no estoy segura de que puede autoevaluarse y que entiende criterios por los cuales tiene que saber autoevaluarse. Que sabe cuándo sabe y cuándo no sabe, cuando tiene que enfrentar un problema real y dice «no tengo suficiente información» o «mi información ya es vieja, necesitaría ver si hay cosas más actualizadas», «esto que hice no está bien, necesito rehacerlo», «no estoy en condiciones de prestar este servicio», «necesito ayuda, necesito asesoramiento, necesito de otro profesional, de otra profesión, porque mis conocimientos en mi propia profesión no me alcanzan hoy».

El tema de la multiprofesionalidad de los problemas es muy fuerte. No podemos encerrar a los alumnos en una carrera, no los podemos encerrar en una asignatura, porque van a tener que trabajar con otras personas. Por tanto que aprendan a autoevaluarse no solamente es un recurso para que entiendan la evaluación y entonces queden conformes, mejoren su aprendizaje, sino que es un conocimiento, una habilidad, una actitud que tienen que tener como graduados universitarios. Y eso lo tienen que aprender a lo largo de la carrera, los profesores tienen que enseñarlo.

Hay maneras de hacerlo. Por ejemplo, al comenzar la enseñanza de una materia les explicamos a los alumnos en nuestro discurso inicial una serie de cosas, qué representa esta materia. Pero también tendríamos que mostrarles cómo son las pruebas de evaluación que tomamos y con qué criterios las calificamos. Tomamos modelos de pruebas que hemos tomado, las fotocopiamos antes de corregirlas y después trabajamos con los alumnos, les explicamos con qué criterios las armamos y con qué criterios las corregimos. No me voy a concentrar en el tema de la calificación, la calificación es lo que sucede después, la evaluación es mucho más que calificar. La calificación no es lo que parece, la escarpela, no es lo más importante. Lo más importante es que el estudiante sea capaz de hacer juicios de valor, con eso trabajamos. Luego tomamos otras pruebas y las repartimos a los alumnos para que las analicen en grupos. Estamos hablando de temas que puedan conocer. A lo mejor les hemos enseñado una unidad y ahora vamos a trabajar sobre esa unidad, para que entiendan con qué criterios estamos trabajando esa materia y qué tipo de aprendizajes esperamos. Repartimos esas pruebas entre distintos grupos, los grupos las evalúan, intercambian las pruebas y las evaluaciones y contrastan los criterios que utilizaron los distintos grupos para evaluarlas.

Este es un proceso gradual, no es que se hace una vez, una vez, una vez, sino que se va trabajando en la construcción de las pautas de aprendizaje de la materia, y de esta manera muchas veces se aconseja trabajar en parejas. Si los alumnos responden a una prueba de evaluación, hacen una autoevaluación y se la pasan a su compañero y su compañero le pasa la de él, intercambian entre ellos los criterios con que han autoevaluado sus pruebas, para llegar finalmente a que

el alumno pueda autoevaluarse. Eso no necesariamente significa que se califique y se autocalifique, sino que pueda analizar cuáles han sido sus dificultades, así, cuando reciba una evaluación formativa de información adicional, en lugar de simplemente mirar la nota y guardarla, probablemente esté en condiciones de usarla. Es decir, tome buenas decisiones de aprendizaje a partir de su evaluación. Es un tema difícil.

Debate

—¿Hay que fijar porcentajes para la aprobación cuando discutimos la escala? ¿Si es 50 %, 60 %, 70 %?

—No puedo contestar eso, porque 50 % ¿de qué?, ¿60 % de qué? Si tomo una prueba de verdadero o falso con 100 ítems, puedo decir que para el alumno apruebe tiene que tener un 80 % aprobado. Pero para eso necesito ese tipo de pruebas, que es una colección de ítems fragmentarios donde cada uno vale por sí, separado de los demás.

Cuando una prueba tiene distintas partes siempre hay dos modalidades posibles de calificación. Una es lo que se llama *calificación analítica*, en la que cada una de las partes vale algo. Por ejemplo: si hay 100 ítems de verdadero/falso, cada uno vale un punto y luego la calificación es proporcional a eso. También puede ser analítica cuando digo, por ejemplo: «Hay que escribir un ensayo, y les voy a poner 40 % por la información que utilizaron, que sea adecuada, que sea completa; 20 % por la redacción; 20 % por la organización del trabajo, y 20 % por la riqueza de las conclusiones». Ahí tengo también analítica, tanto por esto, tanto por esto, tanto por aquello, después sumo y convierto lo analítico en una calificación final.

La otra modalidad es la *calificación holística*, cuando no hay partes que valen cosas separadas sino que la mirada del profesor es global. En las investigaciones sobre este tema tienen mayor validez predictivo las calificaciones holísticas que las analíticas. Los docentes tenemos experiencias, cuando hacemos calificaciones analíticas después nos da algo y decimos «esto qué tiene que ver». De esta manera algunos comenten errores espantosos, pero como la calificación es analítica hay que aprobarlos porque un error no anula todo el resto. Las calificacio-

nes holísticas parecen tener mayor validez, pero son más difíciles de comprender para el alumno y de justificar para el profesor, entonces si no las podemos explicar cuidadosamente la evaluación no se puede resolver bien. En ese caso a veces conviene combinar las dos cosas y decir «voy a hacer una calificación analítica —porque lo que se haga debe ser anunciado previamente al alumno, no puede aparecer después—, por ejemplo, cada una de las preguntas de un cuestionario vale tres puntos, pero si aparecen errores muy importantes no va a estar aprobado aunque la suma dé». Pero hay que advertirlo antes, no puede aparecer después. De todas maneras, mi consejo, que trata de ser coherente con mi idea pedagógica, es decirle al alumno, si ha cometido un error garrafal, «volvé a estudiar, preparalo de nuevo, hacelo de nuevo». El tema no es cómo terminó simplemente, terminó equivocándose, el tema es qué pasa después, y qué pasa después es siempre darle la oportunidad de que corrija, rehaga. Tiene un límite, no puedo decirle «te voy a dar dos oportunidades para que lo rehagas», o «te voy a dar una oportunidad», lo tiene que advertir también. Todo lo que pase en la evaluación tiene que ser dicho previamente, aclarado, respondido, la pregunta siempre tiene que poder ser justificada. Es indispensable.

—No sé si preguntar este tipo de corte. Creo que tiene que ver más con un tipo de aprendizaje, de desempeño esperado. Que ese nivel de desempeño tiene que estar de acuerdo con lo que está, el contenido que está en la evaluación. Si las evaluaciones refieren a niveles de desempeño claros, de criterios establecidos, entonces no importa que esa escala de calificación vaya hasta el 6, hasta el 10 o hasta el 12. En cada nivel tenemos un criterio establecido con determinadas estrategias bien aclaradas, entonces podemos hacer el nivel de corte con su nivel de desempeño, más allá del porcentaje.

—Es como una traducción. Además de las escalas de calificación numéricas hay otras, están las escalas alfabéticas, las escalas conceptuales.

—Es otra forma de decir lo mismo, los niveles del excelente bajan a nosotros con 9-2, los niveles de muy bueno varían de 9 a 12.

—Te voy a decir dónde está la diferencia. La escala numérica que usamos es una escala con nombre de números, pero las escalas numéricas verdaderas son las que se llaman técnicamente *escalas de in-*

tervalos. Son escalas en las cuales las distancias que hay entre pares consecutivos son conceptuales, lo cual permite sacar promedios. Pero si yo no puedo garantizar que la distancia entre 4 y 5 es exactamente igual a la que hay entre 5 y 6, no puedo decir que 5 es el promedio entre 4 y 6. Incluso como nuestras escalas tienen números nos damos el lujo de sumar, dividir, sacar decimales. Es decir, jugamos con los números, pero lo que hay detrás no son números, es cuantificar los saberes artificialmente. Yo puedo medir las aptitudes porque es una escala ad hoc, para medir la aptitud, pero siempre tengo que decir que es relativa a esa prueba, que el alumno aprobó relativamente a esa prueba o no aprobó en relación con esa prueba. Si le hubiera tomado otra, ¿qué pasaría? Lo engañoso son los números.

—Yo soy docente en el Área de Tecnología Médica. Cuando tomo los cursos siempre digo los objetivos de la clase, las formas de evaluar y en consecuencia evaluo de distintas formas, primero los evalúo desde el punto de vista del conocimiento y les pido una monografía, que entiendo que es formativa porque después hablan de esa monografía, pero también los evalúo desde el punto de vista ético, cómo se comprometieron con el trabajo, cómo se relacionan con sus pares, con el docente, con la comunidad. La pregunta es si ese tipo de calificación sirve para todas las materias, porque me pregunto cómo se puede hacer en estadística para evaluar de esa forma cuando después van a salir técnicos que necesitan tener saberes básicos más duros.

—Según el tipo de saberes que uno trata de desarrollar hay que encontrar los modos adecuados de evaluar. Lo que pasa es que tenemos, primero, una escala de calificaciones que es universal, universal en este sentido, para todas las materias exactamente igual. Lo cual implica que si tengo 13 niveles de calidad pienso que en todas las materias puedo diferenciar 13 niveles de calidad, lo que no siempre es fácil de hacer o igualmente fácil. Cualquiera de las calificaciones que ponemos es relativa a las pruebas que construimos, por eso pongo el acento en la prueba, no tanto en la calificación, porque lo que no deberíamos hacer es sujetar las pruebas que tomamos a las formas de calificación que tenemos, porque generalmente es al revés, porque tenemos que calificar, entonces tomamos la prueba de tal manera, con lo cual hay

saberes que te preocupan y que a lo mejor otros dejan afuera. Cada una de las cosas requiere alguna forma de evaluación, y es cierto, si hay saberes duros el alumno tendrá que demostrar que los posee y esa demostración a lo mejor no se hace de la misma manera que en tu materia. Repito, para evaluar necesitamos evidencias de algo, evidencias de que el alumno conoce algo, de que el alumno es capaz de razonar y de inferir cosas, de que asume con responsabilidad una tarea, de que interactúa con los otros y es capaz de escuchar y entender lo que los otros dicen, que es lo que queremos que aprenda. Cada uno de esos saberes se demostrará de manera diferente, la cuestión es encontrar la manera adecuada. Es importante que el alumno entienda que es una manera adecuada para evaluarlo, que lo crea.

—Mencionabas hoy la dependencia que tenemos, al momento de corregir, del orden en que corregimos, y decías que además es incurable. Me voy a atrever a desafiar un poquito eso a partir de la evidencia que tenemos en el sentido de que puede ser muy puntual. Yo estoy convencida de algo que puede ser una norma para todos: que cuando uno se empeña en algo y le deja tiempo, si tiene una preparación sólida detrás, alguna vuelta, algún método siempre encuentra. El problema es que requiere tiempo.

—El método es corregir dos veces, independientemente. Si tenés una pila, corregís así, colocás las calificaciones y después la das vuelta.

—Yo tengo un método un poquito distinto.

—Si me decís que corregís todos los 1, después todos los 2...

—Los ordeno de mayor a menor y los corrijo.

—Esa es una manera, es hacer una primera corrección y después dividir en partes: los mejores, los del medio, los peores.

—Eso es relativo.

—De todas manera algún riesgo se corre, porque es difícil que saques de los bajos a los que pusiste como bajos; los que pusiste arriba no me preocupan, el problema son los de abajo. Estamos viendo qué hacer con el tema de que el orden de la corrección influye en las calificaciones. Yo dije que era incurable, en este sentido de que todas las investigaciones muestran que pasa eso y lo muestran desde hace 100 años. Decía que a pesar de que está publicado, que se conocen los re-

sultados, de todas maneras sigue sucediendo. Eso no quiere decir que no se pueda hacer algo, lo que hace ella es una de las técnicas que se usan. Hay una que es un clásico: las corrijo dos veces con calificaciones independientes cada vez, cuando encuentro que hay una diferencia importante —no de un punto, en ese caso elijo una de las dos—, de más de un punto entre las dos calificaciones puestas independientemente, la vuelvo a leer o se la doy a otra persona. Nosotros solemos hacer eso.

Otra forma que se recomienda es separarlas en tres partes, hacer primero una rápida lectura, colocar en un grupo las que a uno le parece que son las mejores, en otro las que están en el medio, y en otro las peores. Esa es una manera recomendada de hacerla. Hay un riesgo, que es que los que se pusieron en el paquete de los peores, de los más bajos, normalmente se quedan en el paquete de los más bajos, es decir que uno ya está haciendo una especie de corte entre los de abajo, es difícil que uno de los de abajo pase a los del medio. A veces es más fácil que con una lectura cuidadosa uno pase al de abajo.

—Igual, probablemente, el área más delicada es la del medio, la que está en el borde entre pasar o no pasar.

—Ya dije que los que están en torno del puntaje de corte, aprobación/no aprobación, no se diferencian mucho.

—La segunda componente relacionada con eso es que también tenemos alguna experiencia —acá localmente, no a nivel de publicarlo sino de trabajar con una forma sistemática— tratando de dar la misma prueba a dos docentes distintos del equipo. Lo corrigen en forma independiente, no se muestran los puntajes, después cuando está se discute. Trabajando así hemos llegado a formar el equipo de manera que las diferencias en las correcciones independientes representan el 3 %.

—En mi cátedra hacemos eso y cuando aparece un caso en que el primer corrector tiene dudas, lo discutimos entre todos o con otros, pero ya estamos tan armados como equipo que no tenemos.

—Eso lleva a una impresión que deberíamos tener todos en cuenta: hay que trabajar en preparar los equipos para hacer este tipo de cosas.

—Y también hay que trabajar en conjunto para construir las pruebas de evaluación, porque muchas veces lo que pasa es que alguien de la cátedra hace la prueba y entonces los criterios pueden diferir, pero cuando se hace colectivamente hay más tendencia a que haya una especie de

acuerdo. Es lo que pasaba con el caso que comentaba de dos personas para 450 alumnos. Ese no tiene solución. Yo vengo trabajando desde hace tiempo con el tema de la masividad en las cátedras, la numerosidad en las cátedras, y hay dos aspectos: uno tiene que ver con las estrategias de enseñanza y el otro con la organización de la cátedra, y cuando no es suficiente el personal para trabajar no tiene una buena solución. El lugar de llegada es la conversación, la conversación es insustituible.

—El tema de la evaluación de acuerdo al criterio, de acuerdo a la norma.

—En la historia de la evaluación se ha pasado por distintas etapas, que corresponden a teorías diferentes. Aquellos que desarrollaron inicialmente la teoría de la evaluación fueron los conductistas, ellos creían que se podía medir, que lo que se podía medir era lo que se observaba y que la medición debía ser cuantitativa, porque querían hacer de la psicología, en este caso, una ciencia. Una ciencia al estilo positivista, al estilo de las ciencias naturales, y crearon los instrumentos de evaluación, en la que la subjetividad va a quedar afuera, porque admiten una única respuesta correcta. Esas son las pruebas objetivas. Instalaron que las pruebas objetivas eran una manera que no se prestaba a discusión. Pero con el tiempo llegaron a la conclusión de que las pruebas objetivas y esa respuesta correcta podían depender de quién había construido la prueba, por lo tanto que lo mejor era pensar en pruebas que pudieran ser generalizadas, hechas por especialistas, especialistas en las disciplinas, especialistas que no dijeran «esto es correcto» o «aquello es correcto», sino que tuvieran más seguridad en la corrección de la respuesta y que además estuvieran trabajando con especialistas en la construcción de pruebas.

De esta manera se hizo el estilo de pruebas que hoy se utilizan mucho para los exámenes generales, que a veces se hacen en un país, a veces se hacen en una jurisdicción dentro de un país, a veces son los exámenes internacionales al estilo de PISA por ejemplo, de TIMS. Esas son las llamadas *pruebas estandarizadas*, que tomaron como modelo las pruebas psicológicas y los tests y se construyeron de la misma forma. Se utilizaban las pruebas estandarizadas, y la calificación que recibían los alumnos surgía de una escala que acompañaba a la prueba.

Los que habían fabricado la prueba habían hecho un procesamiento estadístico, porque la habían aplicado a una muestra representativa, entonces los resultados estaban previstos y la calificación del alumno correspondía a eso, en función de una comparación que se hacía entre el rendimiento de un alumno ante una prueba determinada —siempre relativo a esa prueba— y otros alumnos.

Pasado un tiempo, como ellos iban revisando sus ideas, llegaron a la conclusión de que había algo que no era correcto, porque si yo tomo la misma prueba a los alumnos de esta aula y a los del aula de al lado, a lo mejor los de acá tuvieron más posibilidades para responder bien a esta prueba algo que se les enseñó que a los de al lado no, por lo tanto su calificación no debiera ser asignada con la misma pauta, o con la misma escala, no era justo. Entonces inventaron los llamados *puntajes derivados*. Los puntajes derivados son por ejemplo la graduación sobre la curva, sobre la curva normal o puntaje Z o puntaje T o puntaje estándar. Son puntajes que se asignan en función de las posibilidades reales que ha tenido cada grupo de alumnos de responder exitosamente a la prueba. La calificación se coloca en dos pasos. Se toma la prueba, se les toma a todos los alumnos en esta aula y en aquella aula, y se produce lo que se llama *puntajes brutos*. Una vez que se han obtenido esos puntajes brutos, se modifican, se procesan en función de las oportunidades reales que tuvo cada grupo de aprender. Para saber cuáles son esas oportunidades reales, cuáles son los indicadores de que han tenido esas oportunidades, se hace un análisis de los puntajes brutos y se obtienen dos medidas estadísticas que son la media y el desvío estándar. Se dibuja la curva y se sabe qué oportunidad tuvieron los que estaban acá y qué oportunidad tuvieron los que estaban allá, y la calificación, que era una calificación bruta, se transforma en un puntaje derivado; es decir, que el alumno va a recibir una calificación según la ubicación que tenga en la curva de su aula, se lo califica comparándolo con sus compañeros. Esto fue una práctica que se adoptó en muchísimos lugares del mundo, en universidades del mundo. Por ejemplo, las secundarias en Buenos Aires utilizan la graduación sobre la curva, es algo que tiene un uso terriblemente difundido, que se sigue usando.

Aproximadamente en 1970, los conductistas, que iban cambiando por generaciones, y siempre estuvieron dispuestos a revisar lo que

hacían, vieron que lo que estaban haciendo no tenía sentido porque el alumno Juan en un curso puede tener una calificación y el alumno Juan en el otro curso con la misma respuesta tendrá otra calificación, en un curso está aprobado y en el otro está desaprobado. Eso no tenía sentido. Vieron que tenían que calificar no comparando los estudiantes con sus compañeros, sino comparando lo que sabían con los objetivos del curso, con lo que deberían saber. La calificación de comparación del desempeño real de los alumnos diferenciados era lo que se llamaba *evaluación referida a normas*. Cuando se adoptan los objetivos como criterio se habla de *evaluación referida a criterios*: sabe o no sabe lo que tenía que aprender. Para eso era necesario formular los objetivos claramente, de tal manera que nadie pudiera dudar acerca de qué es lo que el alumno tenía que saber. Como eran conductistas, esos objetivos debían formularse en términos de conductas observables, cuantificables y debía explicitarse en qué condiciones el alumno tenía que dar esas respuestas, que eran respuestas observables y cuantificables. Es decir, qué tiene que decir, escribir o hacer el alumno, que además tiene que ser visible y medible cuantitativamente. Esos son los *objetivos operacionales*, que representaron algo así como la culminación de toda la corriente conductista en el campo de la evaluación.

Hoy discutimos mucho desde el punto de vista pedagógico, no aceptaríamos los principios del conductismo. Somos constructivistas, este es el enfoque pedagógico que adoptamos, sin embargo los principios conductistas de la evaluación siguen utilizándose. Empecé diciendo que hay incoherencia entre el enfoque pedagógico y el enfoque de evaluación; lo interesante es que en muchos casos el currículo en términos de competencias específicas vuelve a ese tipo de definiciones. Ahora tengo competencias específicas, sé lo que el alumno tiene que hacer, lo tengo que observar haciéndolo y entonces lo puedo calificar. Es lo mismo, aunque se sostengan otras ideas pedagógicas.

—Nos traicionamos en la acción.

—Porque en la práctica es más fácil. No hay peor régimen de promoción que el promedio, no existe, si lo analizáramos cuidadosamente veríamos que no sirve para nada, no es descriptivo. Si un alumno sacó 4 y 10, tiene 7 de promedio, y otro que sacó 7 en los dos también tiene 7 de promedio. No sirve para nada, no describe nada. Para que sea posible un promedio tengo que tener una escala de intervalos, la esca-

la numérica tiene que ser en serio numérica, y no tenemos una regla para medir, no tenemos el metro patrón para medir los saberes de los alumnos. En los promedios se ponen en el mismo nivel de importancia cosas distintas, habría que usar promedios ponderables, trabajos de distinta naturaleza deberían ponderarse de distinta manera, con importancia distinta. Esto sería indispensable. Pero seguimos usando promedios, porque es lo más fácil de calcular. Es relativo.

—En la FEUU nos propusimos avanzar hacia una escala más de corte cualitativo, a una evaluación más de corte cualitativo en este sentido de las calificaciones. Nuestro gran objetivo es una escala conceptual de seis niveles. Pero más allá de la escala, lo que nos interesa es que el estudiante pueda tener una evaluación formativa, que la evaluación sea parte del aprendizaje, ese es el objetivo. Nos encontramos en este proceso con una dificultad muy grande, que es cómo vincularnos con la sociedad en sí, con las instituciones, en distinta forma el sector productivo, porque fuera de la universidad, fuera del lugar donde se trabaja la enseñanza, lo que quieren saber es quién va a hacer de mejor forma el trabajo dependiendo de los niveles de calidad. Es que al estudiante no le importa mucho, es una chance, es poder estar en un nivel de calidad que le permita acceder a determinados grados laborales.

—Grados laborales, becas.

—Creo que es un problema social una necesidad de que me den bien los números, cómo los números estén instalados. Me genera una problemática específica el cómo, reformular ciertos objetivos pedagógicos, cómo los dejamos en las evaluaciones para ver cómo nos ventila la realidad y cómo no, terminar haciéndonos trampa una gran cantidad de estudiantes. Y además en esto hay una gran diferencia en los estudios, hacia dónde se inclinen. Quería transmitir esa problemática.

—No es fácil formar a una persona. Lo primero que te tengo que decir es que una única institución, una sola institución no puede cambiar los sistemas de calificaciones que se usan en el país, y que en una universidad cualquier escala de calificaciones que se utilice tiene que poder ser explicada y tiene que encontrar sus correlatos en esas demandas que estás diciendo. Acá hubo una universidad que cambió la

escala, pero después tuvo que volver a la 0-12, porque su intención fue reducir el número de grados, pero después sus calificaciones no eran comparables con las de otras instituciones, y cuando no son comparables es muy difícil resolverlo. De todas maneras, la Universidad de la República tiene suficiente peso como para hacer los cambios. La que había cambiado la escala era la Católica, y no la pudo sostener, pero la Universidad de la República tiene tanto peso en el país que sí podría cambiarla.

—Hay que cambiar el lugar de aprobación, porque nosotros estamos diferenciados del resto del nivel educativo, tenemos el 3, cuando algunas universidades privadas ya establecieron el 6.

—Pero 3 o 6 es lo mismo.

—En lo conceptual hay una diferencia importante, porque uno es un regular y lo otro es un bueno.

—Los estudiantes tienen que recibir una calificación que sea buena, porque si no aprobó, no aprobó, y si aprobó no es regular, es bueno. En ese sentido el equivalente conceptual tendría que cambiar. De todas maneras hay un desbalance muy grande cuando tenés tres calificaciones para la desaprobación (0, 1 y 2) y diez para la aprobación.

—Prácticamente se demostró que no tenemos niveles de excelencia, porque está todo en el nivel del medio.

—Eso tiene que ver con decisiones que toman los profesores acerca de cómo usar la escala. Pienso que 0-12 es un número excesivo de grados. Si me preguntan, con cinco alcanza, dos para la desaprobación y tres para la aprobación, y habría muchas menos discusiones y discrepancias entre los profesores, pero no siempre se puede hacer esto. Entonces está claro que tienen que ser doce o trece, salvo que se pueda tomar una decisión muy fuerte. Pero las veces que se ha cambiado de 3 a 6, por ejemplo, los profesores que antes pedían 3 ahora piden 6, y cuando es a la inversa lo mismo. Hay que dar una discusión acerca de qué significa. Es difícilísimo. Me ha tocado trabajar en algunas instituciones, incluso escuelas, y es imposible que los profesores se pongan de acuerdo.

—Ella apuntaba a que después en los hechos, cuando los muchachos van a conseguir trabajo, se les pide la escolaridad, y si un R regular en uno es igual a un B en otro, cuando llegan las escolaridades unos

llegan con aprobaciones con R y otros con aprobaciones con B. En la evaluación institucional en distintos lugares aparecen esas cosas.

—No lo pueden aprobar con R, regular es 2, no es 3. Hay que tener cuidado, el equivalente conceptual tiene que ser adecuado. En la UBA había una facultad que por muchos años tenía profesores que ponían solo dos notas, 4 u 8; la nota máxima de aprobación es 10, la nota mínima de aprobación es 4, y los profesores no se comprometían en decir «este es brillante», tenía que ser un genio para que le pusieran 10. Si estaba bien le ponían 4 y si estaba un poco mejor, 8. Habían convertido su escala en otra escala: 2 cuando perdían, 4 cuando estaba bien y 8 para el muy bueno. Era una facultad muy numerosa además, tenía muchos graduados que se iban al exterior a hacer posgrados, entonces empezaron a preguntar, a mí me empezaron a preguntar desde Estados Unidos, porque había alumnos que habían perdido alguna materia y se recibían con 2 de promedio, porque en las que estaban bien sacaban 4 y si habían perdido algunas estaban por debajo del 4. Yo contestaba que alguien que se recibió con 8 de promedio es excelente, alguien que sacó 7 de promedio es excelente. Finalmente hubo una señora norteamericana que ganó una licitación que se hizo en Estados Unidos para sacar un libro sobre qué significan las notas de grado en los países para saber cuando se presenta algún estudiante a continuar estudios del grado o de posgrado.

En el proceso de Bolonia también aparece eso, la explicación de qué significa. Por suerte con los años y no hace demasiado tiempo en esa facultad los profesores también entendieron que no podía ser, primero que la escala era del alumno, porque siempre está esa famosa historia del profesor que decía 10 es para Dios, 9 para el autor del texto, 8 para el profesor, 7 para el mejor alumno, y el 2 es para el alumno, no es para el profesor. Es horrible, pero muchos profesores piensan eso. La escala es del alumno, la escala tiene que representar los aprendizajes de los alumnos, pero acordar con los profesores, tratar de ir acordando el uso de las calificaciones. No es fácil, pero a lo mejor progresivamente se va logrando o se van sumando cátedras.

—Una de las cuestiones que con la escala se complicó más a mi entender en la discusión es que según muchísima gente la forma de

corroborar si las notas están bien puestas es si cuando los docentes terminan de corregir la cantidad de gente es una curva normal sobre la escala.

—Muy buena pregunta.

—Los docentes no afirmaban eso, por eso se discutió en la ATC y fue un estudiante que hizo esa afirmación a partir de la respuesta de la mayoría de los estudiantes, una cosa en la que no estaban de acuerdo. Lo que más se usa es que la masa crítica tiene que estar concentrada en el medio, en la mitad, en la media.

—La idea sobre este tema es una idea filosófico-sociológica. Dada una población cualquiera siempre vamos a encontrar un pequeño grupo de excelencia, buenos, una gran cantidad en el medio y un pequeño grupo de malos. Muchas veces las cantidades son similares y eso es lo que te da una curva en forma de campana. Pensar eso es pensar algo de qué es la sociedad, cómo son los hombres, es pensar que no todos pueden ser buenos, necesariamente tiene que haber un grupo de malos, necesariamente solo puede haber un grupo de buenos. Es una idea horrible. Entre los materiales que traje para el curso que voy a dar mañana hay un artículo sobre la historia del uso de la curva en forma de campana. Está en inglés. Eso está muy instalado en la forma de pensar de la gente y es de un nivel de discriminación horrible. Por eso digo que es el primer referente de qué pienso de la sociedad, qué pienso de las personas, qué pienso del ser humano.

El segundo nivel es referido a las pruebas de evaluación, pensar que una prueba de evaluación es buena cuando diferencia entre las personas. Yo también he escuchado decir que no es buena porque no me da una curva en forma de campana, pero si yo tomara una prueba en la que lo que me interesa saber es si los alumnos saben las cosas básicas, fundamentales, sin las cuales no debieran aprobar mi materia, y todos respondieran bien, ¿qué hago? No digo «esta prueba está mal porque no diferencia entre las personas». Si logro que todos tengan éxito, es el triunfo de la educación, no es el fracaso porque no logró diferenciar los muy buenos de los muy malos. Pero hay técnicas en la construcción de pruebas de evaluación que están basadas en la idea de la discriminación.

Hace unos cuantos años, contaron cómo habían fabricado la primera prueba de evaluación de la calidad del sistema educativo que se hizo en la Argentina, y contaron lo siguiente: habían pedido ítems de pruebas de *multiple choice* de las provincias y les habían mandado una cantidad. Se las dieron a responder a chicos de escuela y eliminaban todos los ítems que todos los chicos habían contestado y todos los ítems que ningún chico había contestado. Se quedaron con los otros, con los que diferenciaban.

A mí me interesa saber lo que todos contestan, lo que ninguno contesta, pero esa idea estaba muy metida en una enorme bibliografía sobre lo que se llama justamente la teoría de Rager, que es la teoría que considera que la prueba es buena cuando discrimina. Tiene tradición, y como tiene tradición está sustentada en una teoría progresista. Es horrible, porque además no es verdadera. Si a la gente se le enseña bien puede aprender, si saben qué tienen que estudiar, si los ayudás a que rehagan, a que resuelvan sus errores, todos pueden aprender. Esa es la cuestión. No van a ser todos genios, pero pueden ser todos buenos graduados universitarios. A ese nivel se puede llegar.

—Hoy decías que la clave era la autoevaluación. Entonces, ¿cómo se genera una cultura para que los estudiantes la acepten? Porque nosotros lo hemos intentado, venimos tratando de que los estudiantes hagan las autoevaluaciones y lo rechazan, no se sienten conformes con el hecho de que uno se la pida cuando hacen una producción. Trabajamos de esa manera y muchas veces se resisten a hacer la autoevaluación.

—En primer lugar, no todo lo podés resolver cuando el alumno ya llegó a la universidad y ya pasó por todo un sistema educativo, tiene toda una historia escolar en la que la evaluación consistía en que otro viniera y dijera. En la medida en esto se vaya introduciendo en los jardines de infantes, en la primaria, en la secundaria, cuando los alumnos lleguen a la universidad la autoevaluación ya va a ser una cosa instalada, algo sobre lo cual han venido trabajando. Para nosotros representa un cambio cultural que es cuando pensás la evaluación como una situación temida. Hay un autor que dice que la evaluación es una situación que genera ansiedad en el alumno e implicación en el profe-

sor. ¿Cómo pasás de eso a pensar que la evaluación es amigable, que te va a ayudar? Es muy difícil.

Y podés encontrar una diferencia en los estudiantes, obviamente encontrás muchísimas diferencias, pero me voy a referir a una que son los niveles de ansiedad. El alumno más ansioso no tolera la autoevaluación, necesita que otro le diga «está bien» o «está mal». Es el alumno que cuando vos tomás un parcial se para veinticinco veces y te dice: «¿Esto está bien?, ¿le parece que está bien por este camino?», mientras que otros cuando les das problemas más definidos, abiertos, se sienten como peces en el agua. El alumno que tiene más resistencia a autoevaluarse es el alumno más ansioso, más dependiente de la evaluación. Eso es difícil de solucionar, hay cuestiones que ya están instaladas en la personalidad, y la realidad es que muchas veces el alumno al que en la casa le exigieron, que va a terminar exigiéndose a sí mismo, se siente sobreexigido, entonces necesita que tu evaluación.

—De todos modos ese objetivo de desarrollar autonomía tendría que poder alcanzarse incluso en esos casos.

—Es indispensable.

—Pero habría que buscar alguna otra receta para la conversación.

—La Universidad de la República, con sus casi 90 000 estudiantes, ¿cómo hace para mejorar los sistemas de evaluación?, ¿cómo se evalúa en la diversidad? Una de las cosas que estuvimos comentando es qué le puedo pedir yo al docente universitario, después de todas estas charlas, de todas estas conversaciones sobre evaluación, cuando sé que no tiene las herramientas, que no tiene tiempo. Sé que estoy evaluando la formación por otra cosa, no por el rendimiento. En los consejos de las facultades se evalúa más por la calidad que por la dedicación en la excelencia del conocimiento. En ese marco tan complejo, ¿qué hago para mejorar la evaluación en esa materia?

—Los cursos masivos te plantean situaciones en las que necesitás recurrir a algún tipo de evaluación que sea práctica. Depende mucho de cómo distribuís las tareas en la cátedra. El tema de la organización no es solamente la elección del formato de evaluación, la construcción del instrumento, sino cómo distribuís las tareas en la cátedra de manera de hacer posible una evaluación más personalizada a pesar de que sean

muchos los alumnos. No siempre es fácil, no siempre es posible. A veces no hay más remedio que recurrir a evaluaciones escritas en las que el alumno tiene que dar respuestas. Mi consejo sería que la construcción sea lo más inteligente posible, que no sean preguntas que admitan una sola respuesta, que requieran el uso de operaciones cognitivas o estrategias cognitivas de orden superior, es decir que le pidan al alumno no que repita sino que razone, que argumente, que haga hipótesis.

Las pruebas de evaluación tienen que ser interesantes y atractivas, y una vez que han sido corregidas tienen que ser analizadas e interpretadas. No basta con tomar la prueba y decir mal, bien, mal, bien, contar, que no es fácil, porque cuando los grupos son muy masivos hasta la suma de los puntos da lugar a equivocación, encontramos muchos casos de errores en esta situación. Algo que nunca se podría dejar de hacer es escuchar al alumno cuando pide aclaración. Eso es obligatorio para el profesor, forma parte de su contrato de trabajo. El alumno ha recibido una calificación, no entiende por qué, hay que conversar con él, explicarle y darle la oportunidad de corregir sus errores. Es decir que puede rendir una prueba, aquellos que cometieron algunos errores deberían tener la oportunidad de rehacerla. No sé si se puede solucionar de la mejor manera, pero se pueden dar muchos pasos para mejorar.

—Al evaluar masivamente, por ejemplo un profesor convaleciente en salud, y yo he estado con tres profesores, siete ayudantes y los alumnos. Mi pregunta es: dado que tenemos que funcionar con esta realidad, ¿cuál es la mejor estrategia para hacerlo lo mejor posible? Hay formas de mejorarlo. Pensar lo mejor posible los ejercicios...

—Primero, no usar un único formato de evaluación, porque cada uno evalúa cosas distintas. Dije discusión inteligente, pero nunca usar solo un formato, si tienen que ser todas escritas porque el grupo es muy numeroso, entonces que alguna parte sea tipo prueba objetiva, que se corrigen más rápidamente —pero tienen que estar muy bien construidas—, y otra parte que sea de producción de desarrollo del alumno, tendría que haber una sabia combinación de las dos cosas. El problema es que el alumno pueda llegar a producir cosas. Acabo de decir que una buena organización y distribución de tareas dentro de la cátedra puede permitir hacer una corrección de ese tipo; algunos se pueden dedicar a

enseñar y otros a las cuestiones vinculadas con la evaluación, la corrección, y en la conversación, la devolución formativa de los estudiantes, ahí se puede hacer una diferencia. Algunos se dedican más a enseñar, otros se dedican más tiempo a evaluar, tienen que ser más frecuentes las evaluaciones, para lo cual tienen que ser prácticas también. Recuerdo casos de grupos muy masivos en los que los profesores resolvieron hacer un gran programa de evaluación, lo más maravilloso del mundo, pero no resistieron dos meses, porque no se puede sostener eso, entonces uno tiene que encontrar una solución práctica.

—Te he escuchado en dos oportunidades, recuerdo cuando dabas ejemplos de clases con más de 700 estudiantes, como pasa en algunos servicios acá, donde hacés una pregunta y el estudiante en un papelito escribe la respuesta, lo que se llama el papel de un minuto o el papel de dos minutos, que no da para calificarlo, es para evaluarse uno y saber qué es lo que tiene que enseñar. ¿Lo usás?

—Son muy eficientes, porque uno rápidamente puede tener una visión de lo que está haciendo con poco esfuerzo. Todas estas estrategias de hace algunos años estaban bien pensadas.

—Es lo que se llama la evaluación CAT.

—Yo soy estudiante de Medicina, de Pediatría. Estamos frente a un dilema. Optamos por esta carrera para ser médicos, y por ahí, por determinadas características, nos hemos dedicado a la docencia, estamos desarrollando algo, pero nadie nos preparó para ser docentes de grado, entonces vamos a intentar trabajar en una cátedra en la que aparece la evaluación. Al comienzo nos ponemos de acuerdo en cómo evaluar y tenemos una instancia en nuestro ciclo frente a un tribunal con paciente verdadero que antes estaba en todos los cursos.

—El caso completo.

—Claro, ahora aparece el ECOE, son actores. Un poco todavía nos negamos porque no tenemos niños actores.

—En Estados Unidos, donde nace el ECOE, utilizan niños actores, y en otros lados también, pero como recibían críticas colocaron a las mamás al lado de los niños de manera de apaciguar a los críticos.

—Como decías, los estudiantes estudian según la forma de evaluar, y cuando llegan a Pediatría estudian pensando que alguien les va a decir, y los niños no les van a decir.

—Pero el ECOE no solo tiene el problema de lo que se llama *pacientes estandarizados* y por lo tanto actores, a mi juicio el problema es el razonamiento médico, porque como consiste en cosas todas fragmentarias que no pertenecen al mismo caso sino que en cada una tiene una competencia diferente, el razonamiento médico no se reproduce. Si se tratara de emergencias podría tener que encontrar una resolución en cuatro o cinco minutos de cosas diferentes instantáneamente, pero si el razonamiento médico implica estudiar el caso, buscar información, encontrar las evidencias, razonar, ver lo que está sucediendo, etcétera, tenés que tener un caso. Pero los defensores del ECOE atacan al caso.

—Nosotros teníamos ese problema con el tribunal. Cuando recién empezamos teníamos un tribunal que era más pasivo y colaborador, sabíamos que con ese paciente con algunos docentes salvábamos y con otros docentes perdíamos. Eso nos pasa hoy, también es una terna. Sabíamos de una evaluación continua del estudiante, una ficha de evaluación que se tiene cuando se empieza el curso con una escala administrativa que es: lo logra, lo logra con ayuda, lo logra sin ayuda; en la instancia con el estudiante se evalúa con el estudiante y generamos la corrección. Es muy gratificante porque pasa de todo, llanto, grito. Y de eso sacamos en la evaluación continua una nota final.

Se hace una evaluación continua consciente y se nota que el alumno crece en determinadas cosas, pero llega al examen final y hay gente que sostiene que esa evaluación es independiente de esa instancia y otros que sostenemos que no. Yo en Medicina evaluo más que el conocimiento, porque creo que el conocimiento lo pueden hacer en otro lado, pero que el docente está para acompañar todo el proceso y hablar más de aptitudes y de otras cosas para evaluar. Y me resisto, me angustia mucho cuando un estudiante mostró compromiso y mostró el problema económico —dicen «yo no sé pero lo voy a usar», en medicina nadie sabe todo—, pero el argumento es que son instancias separadas, que tiene que haber una injerencia. El estudiante tiene una buena relación contigo, pero se presenta al examen final y no le va bien con ese paciente por el temor que le tienen, porque lamentablemente la

evaluación es más punitiva que formativa, se enfrenta por primera vez a un tribunal en el que no tiene peso lo formado por nosotros.

—Algo está mal, porque si la evaluación continua fue bien hecha debiera ser predictiva del examen final. Si el examen final no coincide con la evaluación continua a lo mejor el examen final no está bien hecho. Hay algo ahí que no está bien, no me resulta fácil decirte que si le fue mal en el examen final no importa, no te puedo decir eso. Creo que tenés que revisar la situación, qué pasó con la evaluación continua y qué pasó con el examen final. A lo mejor está mal construido el examen final, a lo mejor a la evaluación continua le faltó algo. En qué casos no coinciden, es probable que no en todos, por qué en algunos pasa eso.

A lo mejor estás omitiendo algo. En nuestra cátedra los alumnos por reglamento tienen que rendir algunos parciales y luego tienen que rendir un examen final. Los parciales son escritos, el examen final es oral, y nosotros encontramos a lo largo de muchos años que cuando los alumnos pasan con la nota mínima (4 o 5) los exámenes parciales, desaprueban el final, las probabilidades son de 90 %. Les avisamos, yo lo indico desde el primer día de clase.

Segunda parte

Las pruebas diagnósticas a debate

Este encuentro se realizó el
31 de octubre 2012
en el Aula Magna de Psicología
(Tristán Narvaja 1674 – Montevideo)
Organizado por:
CSE - Unidad Académica
Red UAE

Mesa redonda integrada por:
Alicia Camilloni
Profesora Emérita de la Universidad de Buenos Aires
Andrés Peri
Director de Evaluación y Estadística de la ANEP
Marina Míguez
Directora de la Unidad de Enseñanza de la Facultad de Ingeniería - Udelar
Modera: Gabriel Kaplún
Director de la Licenciatura en Ciencias de la Comunicación - Udelar

Alicia Camilloni

Para empezar, voy a hacer algunas precisiones referidas a las dificultades que representa la evaluación. Si quiero saber cuál es el estado actual de los conocimientos, de los saberes de los estudiantes o de las personas que se presentan, tendría que diferenciar estados de saberes. En el sentido de que ante una consigna dada, una pregunta dada, el estudiante es capaz de responder o hacer algo. A partir de esa respuesta tengo una evidencia de que ante esa pregunta particular ha respondido o ante esa consigna en particular ha producido algo. Pero

también tendría que considerar qué es lo que le ha permitido responder eso, dar esa respuesta o producir ese producto como sea que este sea. En un caso estoy diciendo que hizo, en el otro caso me pregunto qué puede hacer con lo que sabe y lo que hizo es simplemente una demostración de lo que puede hacer con lo que sabe.

La prueba de evaluación es un instrumento construido para saber qué cosas hace ante determinadas consignas, pero detrás de ellas, creo que con mayor importancia, qué es lo que le permite responder a esas preguntas, a esas consignas de trabajo. Qué conocimientos tiene no es algo que responda a la prueba de evaluación constituida simplemente por preguntas o consignas de trabajo, sino por qué a partir de esas preguntas o consignas. Yo sé lo que él me contesta, pero no sé lo que hay detrás, y no solo no sé lo que hay detrás, sino que no sé qué otras cosas podría contestar, a qué otras preguntas podría responder o a qué otras consignas podría responder, exitosa o no exitosamente.

Tengo que diferenciar lo que es la evidencia observable, sin la cual no puedo hacer una evaluación, de aquello que la fundamenta, que la sostiene, sobre lo cual se apoya esa respuesta que constituye la evidencia. Por lo tanto puedo hacer dos niveles de lectura de las respuestas que da el alumno ante cualquier prueba de evaluación. Ante esta prueba respondió de tal manera; ¿cómo respondería a otras pruebas?, ¿qué conocimiento le permitió dar esta respuesta?, ¿qué significado otorgan algunos a las consignas o propuestas de trabajo que se les dieron? La evaluación básicamente —y esto vale para todas las pruebas de evaluación, no solamente para las diagnósticas— me brinda una evidencia que debe ser interpretada.

Temporalmente, yo pregunto, me responden e interpreto la respuesta en función de lo que la persona está fundamentando, lo que está permitiendo, lo que el conocimiento hace posible. Desde el punto de vista de la evaluación me interesa más saber cuál es el conocimiento que hace posible la respuesta que limitarme a conocerla. Son dos niveles, y de esos dos niveles jerarquizaría sobre todo el segundo nivel, que implica siempre necesariamente interpretación. Pero sería muy simplista hacer este relato, esta descripción del proceso, porque cuando construyo mi pregunta o mi consigna de trabajo para el alumno, ya estoy presuponiendo que a partir de ese trabajo, de esa evidencia que

voy a reconocer, voy a estar en capacidad de saber qué es lo que le hace posible esa respuesta o ese trabajo.

Es decir que es bastante más complejo. Me quiero referir más bien a las razones por las cuales construyo esa prueba, que espero que a su vez me brinde una evidencia que pueda interpretar para determinar qué es lo que sabe el alumno. Por tanto me voy a referir básicamente a ese momento en el cual decido cuál va a ser la prueba, cómo la construyo y por qué, en función de ciertos requisitos que a mi juicio tiene que tener.

En primer lugar, a esa prueba que voy a adoptar la denomino *prueba diagnóstica*. Con eso pretendo dar cuenta de los conocimientos que tiene el alumno y lo hago desde un lugar institucional, estoy dándoles una validez social a la prueba y a las conclusiones que saque de ella. Quiere decir que no estoy administrando sencillamente una ejercitación, un examen para saber algo, sino que institucionalmente, la prueba diagnóstica, en cuanto se relaciona con el ingreso a la universidad, sea vinculante o no, de todas maneras es una representación simbólica de lo que la universidad pretende saber si los alumnos saben.

En este sentido las pruebas que yo he construido con ese fin tienen, adquieren, asumen, un papel social importante, adoptan carácter público y desde el punto de vista institucional me tengo que preocupar por demostrar, fundamentar su validez social. En ese caso tengo que demostrar también por qué se postula de esa manera, por qué se la reconoce con ese valor, por qué está hecha así, por qué se pregunta eso de esa manera. Cuando se adoptan pruebas, por ejemplo diagnósticas, con este carácter público porque son institucionales y asumen su validez, se espera que sean reconocidas socialmente, la prueba adquiere lo que algún autor llama *valor normológico*, yo lo llamo *valor normativo*. La prueba prescribe: esto es lo que se debe saber; repito, no se refiere solo a que tenga respuesta a esta pregunta, sino a que debe poseer el conocimiento necesario para poder responder esta pregunta. No estoy pensando solo en el primer nivel de superficie, sino en un nivel más profundo, en lo que hace posible la respuesta. En ese sentido estoy dándole un valor deseable, estoy prefiriendo que eso es lo que se debe saber, que es necesario. Y puedo avanzar más, por el carácter propio de este tipo de prueba estoy diciendo que es imprescindible, digo que es normativo.

Quando construyo una prueba y tomo decisión respecto de qué voy a preguntar en ella y cómo lo voy a preguntar, por ejemplo, o qué consignas de trabajo voy a dar, estoy diciendo que el estudiante debe responder con ciertos conocimientos que son imprescindibles o que son necesarios, pero que debe demostrar que los posee o que están asociados con conocimientos que son necesarios y que deben tener sentido. Son pruebas que tienen un valor normativo, tienen un valor simbólico, social, público, institucional, y en segundo lugar tienen el carácter de normatividad de lo que se debe responder. Eso ocurre con todas las pruebas de evaluación que utilizamos, por ejemplo, para certificar los conocimientos de los alumnos. Estoy hablando específicamente de las que tienen relación con el ingreso a la universidad, en las cuales se trata de ver cuáles son los conocimientos de entrada de los alumnos. Se considera que esos son los conocimientos debidos, son mensajes, son una aclaración de cuáles son los conocimientos debidos.

En segundo lugar, aparecen algunos de los problemas muy difíciles de resolver en la evaluación, en toda evaluación, cuanto más importantes sean las decisiones que se toman a partir de las pruebas de evaluación, más importantes son los requisitos a los cuales me voy a ir refiriendo. Muchas veces en inglés se hace una distinción entre las pruebas que se utilizan para tomar decisiones muy importantes y las otras pruebas en las cuales los resultados pueden ser sustituidos por otros resultados más recuperatorios, con distintas instancias de evaluación y sobre las cuales no se toman decisiones fundamentales.

De hecho en las pruebas diagnósticas uno podría adoptar decisiones fundamentales cuando se dice que esto tiene tales consecuencias si no se cumplen o no las tienen. En ese sentido, no es una cuestión menor que sean vinculantes o no. Sin embargo, en la medida en que son declaraciones públicas y se hacen tal cual son los conocimientos que se supone que los estudiantes deben tener, se nos plantean con mucha fuerza cuestiones tales como: ¿cuál es la representatividad que tiene el muestreo de contenidos que estamos haciendo en la prueba diagnóstica que estamos construyendo? La representatividad de la muestra no es fácil de resolver, lo que se denomina *validez de contenido* es una de las cuestiones más difíciles de solucionar, yo no puedo tomar todo, nunca se ha podido tomar todo, y en ese caso uno tiende a

recurrir a una selección. Se toma la prueba como que va a constituir un universo significativo de la muestra de contenidos.

No podemos evaluar el conocimiento de todo, por lo tanto recurrimos a los muestreos. ¿Cómo hacemos un muestreo que sea realmente representativo? Si hay algo difícil es hacer eso. ¿Cómo representamos a partir de algunos contenidos el total de los contenidos? ¿Por qué la demostración de que se puede responder o no a determinadas preguntas relacionadas con ciertos contenidos puede ser generalizada? ¿Por qué el alumno que responde —esto se nos plantea todos los días cuando vamos a analizar una prueba de evaluación— a ciertas cosas puede ser aprobado y promovido y un alumno que no responde a esas cosas no? Quizás, ¿hasta qué punto este muestreo, es un muestreo representativo? Tenemos problemas de contenidos, en primer lugar estamos considerando que son legítimos. Vamos a hacer una declaración pública: es lo que hace falta saber y entonces confiamos en que es legítimo que el alumno deba responder correctamente a ciertas preguntas relacionadas con esos contenidos; y también puede ocurrir, que también es muy importante, que no responda. ¿Qué no responda?, ¿qué significa esto?, y vuelvo a decir que me ocupo del segundo nivel y no solamente del nivel superficial de la respuesta a la pregunta.

El muestreo es indispensable, pero no solamente se refiere al contenido de temas de formación, sino que el tipo de trabajo que se hace con los contenidos define los contenidos. Cuando elijo cierto tipo de pregunta, cierto tipo de problemas, cierto tipo de casos para que sean resueltos por el alumno en una prueba de evaluación, estoy definiendo los contenidos; porque no es lo mismo preguntar de una manera que de otra. Las respuestas que se pueden dar a cierto tipo de preguntas formuladas de una manera no se pueden dar si son formuladas de otra. Entonces la representatividad se va a referir no solamente a cuáles son los temas más importantes que elegimos preguntar, sino también a la información del tipo de pregunta, del formato que les voy a dar a las consignas de trabajo o a las preguntas que incluyo en mi primera evaluación.

Qué tareas espero que hagan, repito, también tienen que ser representativas dentro de este muestreo. Las tareas representan procesos cognitivos. Cuando hablo de contenido, estoy hablando de un

tema o una información que debe ser procesada cognitivamente por el estudiante de acuerdo con el tipo de forma con que se le plantea en la pregunta o en la consigna de trabajo. Entonces mi problema para construir la prueba está en seleccionar adecuadamente no solo los contenidos para que representen el todo, sino también el tipo de tarea que voy a pedir, en tanto implica maneras de trabajar esos contenidos de manera de conocerlos. Los procesos cognitivos por lo tanto son fundamentales. Pero esos procesos cognitivos pertenecen al mundo privado de las personas, a su mente. Entonces yo necesito que el alumno frente a esa pregunta o a esa tarea que se le pide que realice demuestre cuál es su desempeño. Es decir que haga evidente ese conocimiento, esos saberes que tiene, esas habilidades que tiene, que se manifiesten para que yo a través de su desempeño pueda inferir cuál es el tipo de conocimiento, el nivel de conocimiento, el significado que está trabajando en relación con esos temas.

De esa manera podemos decir que el muestreo, la representatividad, es aceptable, y tenemos que demostrar hasta qué punto es aceptable; porque si estamos inevitablemente condenados a ese muestreo, ese muestreo puede considerarse significativo porque podemos sacar conclusiones respecto de; este es el tema nuestro, el conocimiento que potencialmente el alumno debiera poder mostrar, porque va a poder mostrar solamente una parte, es parcial el tipo de acercamiento que podemos hacer mediante cualquier tipo de prueba de evaluación, porque consideramos que potencialmente podría cubrir mucho más de lo que le estamos preguntando y con el mismo nivel de calidad que podemos inferir de las respuestas que da.

En resumen: en primer lugar hay un requisito, el de la validez social, que tengo que fundamentar y que tiene carácter formativo. En segundo lugar, esto se tiene que conformar a través de este otro requisito que es la representatividad de la pruebas respecto del total de los saberes que el alumno debiera poseer respecto de ese dominio de ese campo del conocimiento.

En tercer lugar, no le puedo tomar todo, necesariamente tengo que hacer un muestreo. Pero ¿hasta qué punto, a partir de sus respuestas, de sus desempeños en determinados campos frente a la prueba que le estoy planteando, puedo llegar a generalizar un juicio acerca de lo que

el alumno sabe? Este es un problema de diagnóstico, es un problema de la evaluación. Estoy hablando de problemas muy duros dentro de la evaluación, a tal punto que si lo pensáramos mucho, seguramente no nos atreveríamos a tomar un examen. Pero la verdad es que tomamos una prueba de evaluación en cualquier caso, hacemos algunas preguntas, sacamos conclusiones, y no limitamos esas conclusiones a esa prueba, sino que las generalizamos a otras pruebas que el alumno pudiera hacer. Tendemos a generalizar ese resultado para ese alumno, no lo leemos como limitado a esa prueba particular. Lo generalizamos para el futuro, para otras pruebas posibles.

Generalizamos los resultados que se han obtenido en una prueba en relación con un cúmulo de tareas distintas, de saberes que están flechando esas tareas en situaciones que podrán ir apareciendo a lo largo del tiempo. Porque a través de esas pruebas de evaluación intentamos no solo tener validez de contenido, sino también *validez predictiva*, y una prueba diagnóstica precisamente concentra el máximo de los problemas que tienen todas las pruebas de evaluación en términos predictivos. Si hay una prueba que tiene que tener validez de predicción es la prueba diagnóstica. Esa es la razón de ser, la esencia de la prueba diagnóstica. Lo que nos demuestra es representativo no solo de lo que hoy sabe, sino de lo que puede saber en el futuro, de lo que puede hacer en el futuro, porque si no, tomaría otra prueba diagnóstica. Si considero que es relevante averiguar cuáles son los saberes que hoy posibilitan su respuesta es porque estoy apuntando hacia el futuro. La flecha hacia el futuro es propia de la prueba diagnóstica; es algo difícil de lograr pero que hay que probar.

No siempre en evaluación uno puede comprobar las hipótesis que ha hecho, pero en el caso de las pruebas diagnósticas sí lo puede hacer, se trata de ver si las predicciones que resultaban de los resultados, de la interpretación de los resultados, se confirman, se pueden corroborar en el futuro. La generalizabilidad de los resultados de las pruebas diagnósticas se refiere más bien a las situaciones futuras y a las predicciones, a las conjeturas acerca de qué podrá ocurrir con los aprendizajes de ese alumno en el futuro, e incluyo algo que considero fundamental: si el alumno por ejemplo no lo dominara, esto sería un obstáculo para un aprendizaje futuro.

Si el alumno no responde a todas estas preguntas que le he hecho, esto es una predicción respecto de que tendrá dificultades en el futuro, eso se puede comprobar. En este caso puedo hacer predicción en relación con evaluaciones futuras y tengo que tener en cuenta cuál es el papel de aquel que es evaluado y de quienes son los evaluadores. No los puedo dejar fuera de estos procesos de interpretación y de producción de inferencias a partir de los resultados de la prueba.

Otro problema que es muy difícil de definir, pero es un requisito y hay que tenerlo en consideración, es el nivel de dificultad que uno le va a asignar a la prueba, cuando consideramos con qué nivel de calidad esperamos que las respuestas se den, pero además cuál es el nivel de dificultades, las preguntas que vamos a hacer, porque en cualquiera de los terrenos tenemos niveles más bajos, niveles medios y altos de dificultades. No digo que sea imposible, ni siquiera una decisión caprichosa, pero es absolutamente indispensable considerar esa cuestión, dónde pone uno ese nivel de dificultad, porque según sea más alto o más bajo, esto generará éxito o fracaso en los estudiantes, y tener éxito o fracaso no es una cosa menor. Si yo estuviera, por ejemplo, trabajando en relación con conocimientos necesarios para el ingreso a la universidad, si me coloco en una posición en la que el alumno comienza sus estudios universitarios, aun cuando no pueda vincular la prueba con una experiencia de fracaso —las experiencias de fracaso en este tipo de situaciones muchas veces se unen a experiencias de fracaso anteriores, a veces son las viejas experiencias de fracaso que tenían los estudiantes—, en ese caso también define expectativas, aspiraciones para los estudiantes, y puede ser riesgosa, por no de decir peligrosa. Por eso hay que definir con muchísimo cuidado cuáles son los niveles de dificultad aceptables. No hay una receta para esto, es una decisión que hay que tomar y que debe estar fundamentada.

Otro aspecto importante es que en una prueba yo tengo un conjunto de ítems, estímulos diferentes, y me tengo que preguntar cómo tengo que tomarlos, o si son cosas fragmentarias y sueltas. No es lo mismo hacerlo de una manera que de otra, no es lo mismo que conforme un todo orgánico coherente en el que quizás surja algo determinado de quien está resolviendo la prueba o quizás tome cosas diferentes. Hay una etapa para cada una de las dos cosas. Cuando son

coherentes tengo ese perfil, cuando no lo son puede ser que tenga una mayor variedad en el muestreo y por lo tanto pueda lograr mayor representatividad, pero al mismo tiempo es más difícil hacer las inferencias predictivas respecto al comportamiento o al desempeño futuro del alumno. El nivel de dificultad es una decisión que hay que fundamentar. Por qué se espera que la forma responda en ese tipo de dificultad, también respecto de que la prueba sea coherente, que sea posible componer un perfil total del alumno o dejarlos como planteos separados de manera que al averiguar si el alumno se desempeña de una manera u otra en distintos aspectos. Quizás haya que hacer una combinación de las dos cosas.

Siempre van a quedar dos requisitos. Un requisito fundamental es que esos conocimientos que el alumno demuestre poseer sean transferibles, porque si están encapsulados, y por lo tanto no son transferibles a nuevas situaciones, tienen un determinado carácter y nivel de calidad, y en ese caso creo que no son predictivos. Si podemos demostrar que son transferibles, es decir que el alumno es capaz de utilizar eso que me muestra en esa determinada prueba en nuevas situaciones, en la resolución de nuevos problemas y de diferentes situaciones y problemas, podemos decir que el aprendizaje que demuestra sí es de buena calidad y podría tener mayor valor predictivo que si estuviera encapsulado.

Otro aspecto típico de este tipo de pruebas es el grado de discriminación. Las pruebas pueden tener objetivos diferentes, en los exámenes generales pueden estar realizadas para ver lo que todos saben o lo que ninguno sabe, o pueden estar interesadas en discernir entre los alumnos para saber quiénes son los que más saben y quiénes los que menos saben. Son dos clases de pruebas totalmente diferentes. En un caso yo pido que no haya discriminación, porque si todos lo saben lo único que sé es que todos lo saben; si ninguno sabe, entonces tampoco los puedo diferenciar porque sé que ninguno sabe. En el primer caso puedo decidir que no necesito enseñar algo porque todos lo saben, en el segundo caso puedo decidir que tengo que enseñarlo porque ninguno lo sabe. Pero los exámenes en general han dado como resultado un panorama totalmente diferente.

Hay técnicas para diferenciar en los exámenes si los grados de discriminación son altos o no. Si yo pensara en una prueba diagnóstica, son más interesantes las que tienen baja discriminación, porque me muestran un camino, si todos saben tengo una plataforma segura, lo que ninguno sabe es lo que tengo que enseñar, y tengo que demostrar que unos y otros son conocimientos importantes y representativos. Decidir cuál es la capacidad de discriminación entre los que más saben y los que menos saben es algo fundamental en la construcción de este tipo de pruebas, depende de cuáles son los propósitos que persigo cuando las construyo.

La validez social de este tipo de pruebas depende del uso social que yo haga de los resultados. Los usos sociales pueden ser «unisex»: entro en la universidad o soy un fracasado. «Soy capaz o no vale la pena que siga adelante» también tiene un uso social, marcar a la gente con éxito fracaso es un uso social. Por eso estas pruebas nos dan una información que de otra manera no tendríamos.

No estoy diciendo que no haya que usar la prueba diagnóstica, tenemos que cuidar su construcción y el uso que le damos, qué, cuáles podrían ser mejores, quizás con los resultados sociales. En ese caso hay un nivel de discriminación, va a favorecer alumnos, a aquellos que les vaya muy bien en la prueba, y evidentemente va a perjudicar a otros, a aquellos que por distintas razones pueden no tener buenos resultados en las pruebas; esto pertenece al mundo del alumno, al mundo de las aspiraciones de cada uno, y es muy difícil de manejar en el caso de los fracasos. Por eso la validez social que le doy a la prueba va a depender del uso social, personal, de las consecuencias que tenga para cada uno de los estudiantes. Entonces son necesarias, me dan una información que de otra manera no tendría, pero hay maneras diferentes de realizarlas.

Andrés Peri

Voy a comentar qué hacemos en la ANEP. No voy a hablar tanto de evaluaciones diagnósticas, pero sí de evaluaciones formativas y de la conformación de un sistema de evaluación de aprendizaje formativo y en línea con dos características: en línea porque usa el Plan Ceibal como plataforma para evaluar, y formativo por el objetivo que persigue.

Esto surgió de un acuerdo de inspectores en 2006 en el que iba a presentar una evaluación regional y la Inspectora Técnica me invitó a ver los resultados de lo que hacía Primaria. Primaria tenía un sistema complejo, donde la Inspección Técnica mandaba una planilla en la que cada maestro reportaba al director el porcentaje de niños con nivel suficiente en distintos dominios como lengua, matemática, ciencias; luego el director consolidaba la escuela y le mandaba al inspector zonal, que consolidaba el distrito y lo mandaba al inspector departamental. El departamental consolidaba el departamento, lo mandaba al inspector regional y el inspector regional lo mandaba al inspector técnico. En esa reunión se presentaba ese largo proceso de construcción y había un cuadro que mostraba suficiencia por grado y contexto sociocultural y daba plana la línea, en los dos sentidos, tanto en grado como por contexto sociocultural de la escuela.

Empezamos a tratar de entender qué quería decir esto y los inspectores llegaron a la conclusión de que sin un instrumento y sin definición de suficiencia explícita, no se podía sumar. No podíamos seguir con este sistema que además de complejo y tortuoso, insumía mucho trabajo. Propusimos llevar adelante en conjunto una primera experiencia en lectura y escritura en primer año. ¿Por qué primer año? Porque en nuestra cultura escolar estaba arraigado de que si el niño no sabe leer ni escribir en primero no puede pasar a segundo. ¿Qué es leer y escribir en primer año? ¿Qué tipo de producción? ¿Qué tipo de comprensión lectora?

Hicimos una evaluación para hacer emerger esos niveles y fue una experiencia interesantísima que luego fue parte del Plan de Lectura que impulsó Primaria. Decidimos seguir con segundo año, y no solamente evaluar lectura y escritura, sino también matemáticas y ciencias. Cuando elevé el presupuesto: 50.000 niños en el grado escolar por tres áreas para evaluar, por 10 hojas por prueba eran 1.500.000 hojas. En presupuesto dijeron que no había dinero para imprimir. Ahí estaba empezando el Plan Ceibal y nuestro ingeniero dijo: ¿Y si no imprimimos? y empezó un desarrollo para transformar el tipo de evaluación que hacíamos, para hacerla en línea.

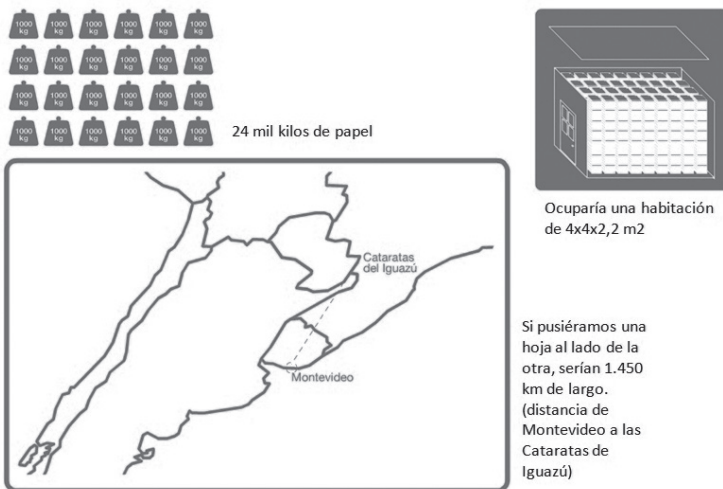
A partir del año 2009 hicimos la evaluación en segundo año en lengua, matemática y ciencias en línea, 35.000 escolares hicieron una

Comisión Sectorial
de Enseñanza

evaluación y los maestros tenían los resultados al instante. Este es el otro gran cambio conceptual. Antes demorábamos cuatro meses en devolver los resultados; ahora el maestro tiene los resultados cuando termina el último niño, el director tiene un consolidado de su escuela y después la DIEE hace un informe general. Ahora nos concentramos en los grados tercero, cuarto, quinto y sexto, y la idea es cubrir los grados de primero a tercero del ciclo básico.

Para tener una magnitud del cambio, si la evaluación del 2012 la hubiéramos hecho en papel habría insumido 24 000 kilos de papel. Si hubiéramos puesto una hoja al lado de la otra hubiéramos salido de Montevideo y llegado a las Cataratas del Iguazú con la cantidad de papel que habría insumido el último ciclo de evaluación. Habría ocupado una habitación de 4 por 4 por 2, el ascensor que tenemos en la oficina habría colapsado. En suma, no podríamos hacer esta escala de evaluaciones si no hubiera mediado una transformación tecnológica que nos permitiese hacerlo.

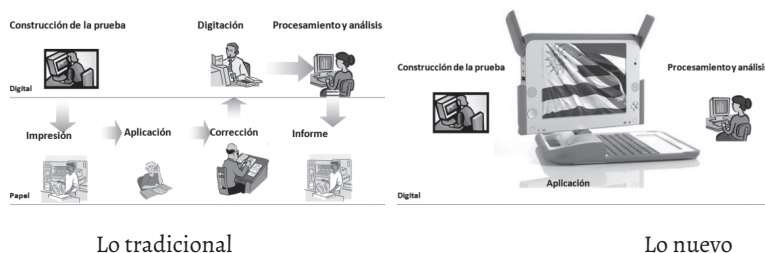
Figura 1. Estimación del gasto de papel del ciclo de evaluación 2012



Empezamos en 2009 con 91.000 evaluaciones, porque son 35.000 alumnos que hacen tres evaluaciones (lectura, Matemática y Ciencias),

por eso se multiplica aproximadamente por tres; y en el último ciclo fueron 540.000 evaluaciones, porque fueron 185.000 los estudiantes que participaron de 3° a 6° en escuelas públicas y privadas.

Figura 2. Esquema de los diferentes modalidades de evaluación papel-computadora



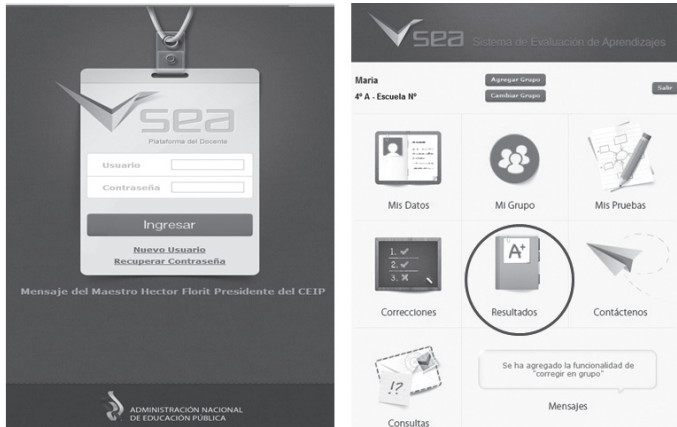
¿Cómo es el proceso si hiciéramos una línea digital y una evaluación en línea? El diagrama que muestra el esquema tradicional empieza por un docente que propone una actividad de evaluación, generalmente desde su computadora, después se conforma la prueba y se manda imprimir. Después de que se imprime, se aplica al papel, se corrigen las preguntas abiertas en papel, después vuelve al mundo digital cuando se digitan o escanean las formas de evaluación, para procesar toda la información, para después volver al papel cuando se imprime el reporte.

En la entrada y la salida del mundo digital insumía casi dos tercios del costo de las evaluaciones. Esto era así hasta que con el Plan Ceibal aparecieron las ceibalitas; ahora podemos ir desde la construcción de la prueba hasta la aplicación, el procesamiento y el análisis sin salir del mundo digital. Este fue un cambio conceptual-tecnológico que nos permitió dar este salto cualitativo.

¿Cómo funciona una evaluación en línea? Funciona en base a distintos los portales. Está el portal de los docentes, el portal de los alumnos y el portal del director. En la pantalla principal del portal de docentes (ver figura 4), aparecen distintos íconos, donde en la sección «Mis datos» aparecen los datos del docente, en «Mi grupo» los grupos que tiene a cargo, en «Mis pruebas» las pruebas que se confeccionaron para aplicar. Cuando llega el día, el docente habilita una prueba, los estudiantes

Al finalizar la evaluación, el docente ingresa a la plataforma, va a resultados y allí le aparecen dos formas de visualizarlos.

Figura 4. Visualización de la Plataforma Docente
(<http://docentes.sea.edu.uy>)



En su bandeja de «resultados» aparecen todas las aplicaciones que se hicieron, donde se muestra cuándo se inició, cuándo se terminó y cuántos alumnos realizaron las evaluaciones. Cuando el docente selecciona el ícono de resultados y tiene dos formas de visualizarlos. Una es la matriz general del grupo, donde le aparecen las actividades de evaluación y le aparece los estudiantes de su grupo. Ahí puede analizar patrones de repuesta. Por ejemplo, cómo le fue a «Valentina» en la prueba, en la lectura horizontal, pero también puede ver qué pasó con una actividad en particular. De hecho si selecciona una actividad, le aparece cuál es la actividad, que competencia y contenido evalúa. A su vez tiene a su disposición una serie de filtros, donde puede elegir que no le muestre todo, y ver, por ejemplo, solo aquellas actividades en matemáticas de un dominio conceptual o una competencia evaluada. Va filtrando y analizando, qué le pasó a su grupo con respecto a las propuestas de evaluación. En resumidas cuentas, la matriz general del grupo constituye una visión panorámica de lo que sucedió en una aplicación concreta; al realizar la mirada vertical puede darse cuen-

ta de cómo se desempeñó su grupo en cada una de las actividades de evaluación propuestas, y en una mirada horizontal puede observar el desempeño individual de un niño o niña.

Figura 5. Visualización de la matriz general

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24			
Valentina	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	18	0.0	
Giuliana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	16	0.0
Marina	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	13	0.0
Marina	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12	0.0
Nicole	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15	0.0
Lucia	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	6	0.0
Facundo	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18	0.0
Mateo	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14	0.0
Iuli	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19	0.0
Mateo	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12	0.0
Facundo	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	16	0.0
Emanuel	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14	0.0
Juan Pedro	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11	0.0
Lucas	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	16	0.0
Gonzalo	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	0.0
Juan Manuel	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14	0.0
Facundo	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11	0.0
Nahuel	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14	0.0
	12	16	0	15	15	13	9	4	8	14	9	14	15	10	8	15	3	13	16	5	15	0	13	6			

También hay una forma de ver la matriz abreviada, por ejemplo, uno puede filtrar y en vez de poner todas las actividades, ver solo las actividades de numeración; en lectura, solamente la clase de texto narrativo; en ciencias solo un dominio conceptual de los dominios que valen más en ciencias. Así uno puede a su vez seleccionar los dominios del contenido. Son funcionalidades para ayudar al docente en el análisis de los resultados. Todo el análisis es autónomo por parte del docente.

Otra forma de analizar la aplicación concreta es actividad por actividad: uno selecciona otra parte de la plataforma, aparece la propuesta de evaluación y elige analizar una en particular. Acá la plataforma muestra una distribución de todos los alumnos de su clase según las cuatro opciones de respuesta. Es importante analizar no solamente los que contestaron bien, sino cuáles son los errores persistentes. Cada una de las opciones fue elegida porque hay una hipótesis de trabajo de un posible error y eso está mostrado en la justificación de cada una de las seleccio-

nes. Eso permite hacer un análisis de por qué mis alumnos contestaron de esta forma y seguir profundizando sobre fuentes de error.

En el análisis por actividad, se muestra una gráfica con la distribución del porcentaje de respuestas correctas en un grupo. Otra gráfica muestra cómo se distribuye el grupo según las cuatro opciones de respuesta. Es importante no solo analizar porcentajes de respuestas correctas, sino por qué opciones optó cada uno de los estudiantes. Uno hace clic en «Valentina» por ejemplo, y aparece la prueba completa de la estudiante ya corregida. Además hay una opción que es la «corrección de actividades en grupo». Esta es una evaluación formativa que se hace a mitad de año y en la que no es crucial el resultado —porque no hay una suficiencia establecida—, sino un referente sobre el cual retroalimentar una discusión pedagógica sobre cómo algunos aspectos fundamentales del programa. Frente a algunas de las cuestiones que planteaba Alicia, la discusión de exigencia y la discusión de representatividad de contenidos es una discusión difícil, más aún cuando tenemos programas por contenidos que no definen expectativas de logros, nivel de exigencia para el alumno en una actividad a mitad de año.

En algún caso es posible que no hayamos acertado al nivel de exigencia. Pero lo realmente importante es que construyó un referente sobre el cual se centra una discusión pedagógica. Lo que importa es tener un referente y que la actividad sea significativa, esté alineada con el programa y responda a un consenso de lo que se espera en la disciplina y en el grado determinado. Ese es un proceso complejo de construcción de consensos que no se ha plasmado completamente. Pero piensen que todo esto surgió a partir de transformar un método tradicional, que era el reporte del porcentaje de niños suficientes. Esto cambia radicalmente el tipo de partido, de referente y de visión de cómo se viene trabajando.

Para profundizar sobre los porqués en una evaluación formativa no importa salvar, lo central es analizar a *posteriori*. Al ser en un medio digital, uno lo hace las veces que quiere, y cada aplicación queda guardada en la plataforma. De hecho, muchísimos docentes dijeron «esto se hizo demasiado pronto, lo voy a hacer en octubre»; es una posibilidad que brindan los medios tecnológicos. Antes, en el formato papel era imposible. Yo siempre digo que si me planteaban evaluar tres áreas

de conocimiento para cuatro grados escolares en papel, renunciaba. Porque la logística asociada al papel es imposible de manejar, a esta escala casi censal.

¿Dónde están las innovaciones?

a) Construcción del Referente

La primera cuestión es una concepción de validez, una edificación de un referente conceptual bien identificado. Cuando se evalúa matemática, por ejemplo, se construye una tabla de especificaciones que define y jerarquiza los contenidos y competencias que se van a evaluar. Después las actividades deben dar cuenta de cada una de las celdas identificadas en la tabla de especificaciones. Y hay un cuidado en la construcción de validez del constructo para poder hacer inferencias sobre los dominios evaluados.

Hemos visto una tensión con los maestros. Lo que ellos más valoran es una actividad concreta, ver qué pasa con esa actividad, y nosotros no podemos decir: «En nuestras evaluaciones en ciencias pasa esto»; porque supone una construcción para arriba de referentes que el constructo ciencias está representado por esas actividades. Por eso, no sumamos ni definimos los mismos niveles de desempeño, mostramos la matriz en forma transparente para que haya reflexión pedagógica sobre qué es lo que pasa.

b) Salto tecnológico

La tecnología no debe ser el eje de los cambios, pero la capacidad de realizar una evaluación a nivel nacional con lógica autónoma y sin enviar un papel ayuda y cambió las reglas con que veníamos operando. Me acuerdo de la anécdota de cuando fuimos a Colonia para ver si se podían hacer estas evaluaciones con las «ceibalitas» en el 2008. Recuerdo que la maestra que nos recibió —creo que era en Colonia Valdense— nos dijo: «¡Y justo ahora vienen con esto, que el año que viene me jubilo!», cuando vio que la evaluación se corregía automáticamente. Pero el proceso de reflexión de qué pasó para obtener ese resultado no lo hace ninguna máquina. Es fundamental para comprender la práctica pedagógica, y eso no está estandarizado. Com-

prender qué pasó es un producto que está en manos de los docentes y lo central en este caso es el diálogo entre docentes de distintas aulas. Es muy distinto en una sala de reflexión cuando dicen: «¿Cómo les fue a los tuyos en la actividad esa sobre los triángulos?». Las actividades entran al debate pedagógico con un referente claro, permitiendo profundizar sobre lo que se espera de la enseñanza y problematizando el aprendizaje de los estudiantes en tiempo real. Este es un aspecto central que potenció la evaluación en línea: la reflexión continua sobre las prácticas.

c) Aprendizaje Institucional

La evaluación en línea permite brindar información significativa a cada nivel. El maestro quiere ver los resultados de su grupo, el director de su escuela y después a nivel general hacemos un reporte para todos los inspectores. Esto supone una nueva gestión de la información, donde ésta no sube ni baja, sino que se genera una base de datos única con accesos significativos para los distintos actores. También supone un cambio en la estructura de costos: si bien no hay que imprimir, hay que arrendar una nube de servidores para habilitar que en un día se hagan 40.000 evaluaciones como esta. Es más barato que imprimir, muchísimo más barato una vez generada la red de conectividad y distribuidas las computadoras.

Me parece que la palanca de los cambios es el aprendizaje institucional, reivindicar el aprendizaje institucional, reivindicar el valor formativo, la posibilidad de retroalimentar en tiempo real, combinando la mirada disciplinar de especialistas sobre la disciplina, la mirada didáctica de los formadores y una mirada técnica de los evaluadores con la orientación de los supervisores: unir en una visión común sobre el aprendizaje. La capacidad de generar un ámbito de discusión local y nacional frente a las propuestas de evaluación —con sus alcances y sus limitaciones— permite pasar de la mera devolución de resultados estadísticos, a la reflexionar sobre las prácticas, para permitir intervenciones pedagógicas oportunas en los centros escolares.

Se debe romper el aislamiento del aula, convirtiendo al docente en una comunidad profesional preocupada por buscar entre todos, la mejor forma de enseñar en aras de lograr un aprendizaje significativo.

A lo largo del tiempo nos fuimos dando cuenta de que debíamos hacer las evaluaciones antes. La primera la hicimos en octubre, la última en junio y julio. Pero cuanto antes uno las hace, menor es la posibilidad de alinear la evaluación a los objetivos de logro por grado. Pero es mayor la oportunidad de trabajar durante el año y sobre esas tensiones trabajamos.

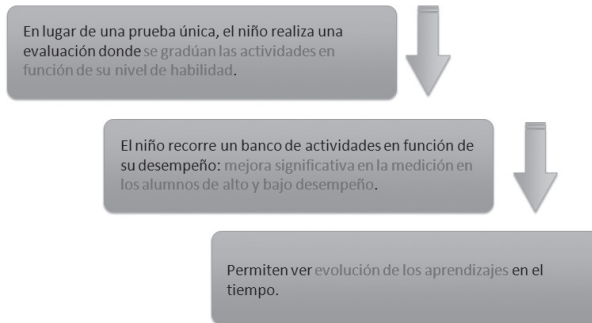
Proyecciones

En cuanto a las proyecciones hacia el futuro, el desafío de personalizar la enseñanza, parece un poco paradójico hablar de personalizar con tests estandarizados. Tenemos un convenio con la Facultad de Psicología, y con un equipo de matemáticos estamos desarrollando tests adaptativos. En los test adaptativos no hay 24 actividades de evaluación para evaluar, sino que según el niño va contestando, el sistema le propone actividades más fáciles o más difíciles hasta determinar su nivel de habilidad. Eso ya lo probamos, ya funciona para la evaluación de matemática del 2009, y la próxima semana empieza un piloto en el que van a participar 26.000 niños y se van a probar 2.200 actividades de evaluación para tercero, cuarto, quinto y sexto, en lengua, matemáticas y ciencias.

Si vemos desde que se empezaron a hacer evaluaciones en el año 1996, partimos de la Teoría Clásica de los Tests, donde se podía comparar en el tiempo, si se contaba con pruebas equivalentes y éstas se aplicaban al otro año. Así se podía analizar qué pasó, si había mejorado el aprendizaje o no. Eso fue así hasta que se nos acabaron las pruebas equivalentes, y tuvimos que cambiar de metodología y pasamos a la Teoría Repuesta al Ítem para evaluar un dominio conceptual. No utilizamos 24 actividades sino aproximadamente 100 actividades por área evaluada. Aunque utilizamos un diseño matricial donde los niños siguen respondiendo 24 actividades pero en cuadernillos enlazados por ciertos bloques de actividades. Esto nos permite evaluar más dominios

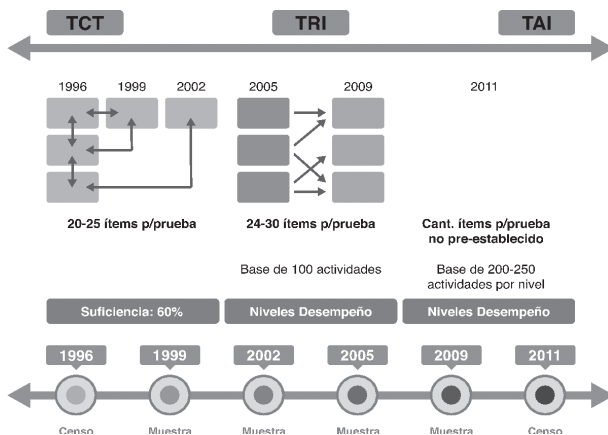
conceptuales que antes, ya que el nivel de suficiencia en la Teoría Clásica de los Tests se determinaba con el 60 % correcto de las actividades de la prueba.

Figura 6. Esquema sobre los Tests adaptativos



La gran virtud de la teoría de respuesta al ítem es que describe niveles de desempeño: estos alumnos pueden hacer esto, estos otros pueden además de hacer esto resolver estas actividades de mayor nivel de complejidad, etc. Se describe una escala de desempeño que es muy importante para la devolución de resultados. El salto que estamos dando ahora es para los tests adaptativos informatizados, en los que la equiparación ya está dada por la aplicación del test.

Figura 7. Inclusión de Tests adaptativos



La potencialidad del test es que me da referencias válidas sobre la potencialidad de los estudiantes durante el año lectivo, permite identificar la dispersión de desempeños de un grupo, contribuye a personalizar la enseñanza y a identificar niveles de desempeño de cada niño, lo que ayuda a la construcción de equidad social al pretender que en todas las escuelas se propongan metas de mejora en el progreso de los niveles de aprendizaje.

El desafío es conceptualizar el lugar que ocupa esta evaluación en el sistema educativo. Más que una evaluación para el carné de notas o para definir promoción, tiene otro estatus, pero es difícil introducir esta nueva conceptualización de una evaluación formativa a gran escala. Por ejemplo, al final del ciclo les preguntamos a los docentes: «¿Considera que la evaluación en línea puede contribuir a la mejora de las prácticas de enseñanza?». Esta es la distribución (<www.anep.edu.uy/sea>). Si uno hace clic arriba de cada una de las barras de la gráfica puede ver las respuestas abiertas de los docentes a la pregunta «¿Por qué?». Por ejemplo un docente que contestó «muy de acuerdo» contestó: «Porque brinda estrategias que se pueden tomar, y abordar los contenidos en forma diferente. Sirve, también, para ir controlando nuestro trabajo. Ver por dónde van las estrategias y los recursos que se utilizan». Otro docente, de los que contestó la encuesta comentó

«Porque presenta un formato nuevo a los docentes, que puede pasar a formar parte de los recursos con los que a diario cuenta la clase». Hay 1.500 respuestas individuales. Si quiere ver las respuestas «muy en desacuerdo», hace clic y las ve. Están todas las respuestas colgadas en el sitio web.

Para terminar, se puede afirmar que es una innovación en fase de consolidación; es difícil decir cuándo una innovación se consolida, pero en el caso de la evaluación en línea cada año hacemos más evaluaciones y cada año nos planteamos un desafío mayor.

Creo que lo central es construir una progresión conceptual en una tabla de especificaciones que trascienda el nivel primaria y medio, ese es el desafío en que estamos metidos ahora, de desarrollar un sistema de evaluación de aprendizaje formativo y en línea de tercero a tercero. Esto es muy importante ahora, para la última parte, en la que yo voy a hacer tres reflexiones más sobre el sistema educativo, porque en las caracterizaciones que hacemos de nuestro sistema educativo solemos decir que tenemos problemas en las transiciones. Tenemos un problema en la transición de inicial a primaria, tenemos un problema enorme en la transición de primaria a media, tenemos un problema enorme en la transición de media a la universidad.

Acá voy a contar dos anécdotas chiquititas. Recuerdo una vez que Pedro Montt —especialista chileno que vino a dar una conferencia invitado por Unicef—, cuando empezó a ver las cifras de Uruguay no lo podía creer. Hace poco vinieron unos amigos míos de Argentina a un congreso que se hizo en Montevideo y decían: «Nosotros revisábamos las cifras de las encuestas de hogares de Uruguay sobre niveles educativos de la población y decíamos: le erramos al programa, no puede ser que Uruguay tenga estas cifras». Porque en nuestro sistema educativo 96 % va a terminar primaria, dos tercios van a terminar el ciclo básico, pero solo un 37 % va a estar en condiciones de seguir la universidad. Es un sistema —el medio— que selecciona tempranamente. De hecho el especialista chileno dijo en la conferencia: «Ya sé por qué ustedes no tienen examen de ingreso a la universidad: porque seleccionan antes».

Y ahí viene una cuestión, en esta discusión de las pruebas diagnósticas tengo la percepción de que hay dos tipos de decanos. Unos decanos que dicen: «necesito que los alumnos que llegan a mi facultad

sepan esto y esto y esto» —muy específico, de la disciplina. Otros decanos que dicen: «Yo lo que necesito es que los muchachos que llegan a mi facultad sean buenos lectores, tengan pensamiento crítico, que yo el conocimiento disciplinar se los enseñe». En esa tensión, nosotros en la ANEP a veces la recibimos como «mirá cómo me los mandan». El «mirá cómo me los mandan» tiene efecto de reverbero para abajo, porque media dice «mirá cómo los manda primaria», primaria dice «pero mirá cómo me los manda inicial» e inicial dice «mirá cómo me los manda la familia». Independientemente del juego de la culpa, que no lleva a solución alguna, lo crítico en un sistema educativo integrado es tener señales de progreso a lo largo del ciclo. Nosotros tenemos un sistema muy fragmentado, por lo cual es un desafío grande desarrollar evaluaciones que trasciendan los ciclos educativos.

Las evaluaciones son una señal y es bueno tener referencias para el sistema. Nosotros no hemos podido construir señales certeras de qué esperamos en distintos niveles de aprendizaje de forma integral, eso es parte de la discusión. Hay una discusión sobre fomentar el desarrollo profesional en las instituciones mediante instancias de reflexión colectiva; yo creo que lo fundamental en algunos casos de las instituciones primarias y medias es construir un referente y aportar a una sala de reflexión. Si la evaluación logra eso, «tiramos cohetes». No tiene que determinar niveles de suficiencia; nuestro desarrollo curricular nos impide avanzar a veces en algunas de esas cuestiones. Es difícil acordar un trayecto común de construcción conceptual de qué esperamos en comprensión lectora, en ciencias o en matemáticas entre inspectores de Primaria, UTU y Secundaria, —aunque hemos hecho grandes avances. Esos son los desafíos de un sistema integrado que necesitamos desarrollar para tener posibilidades de mejora continua.

Marina Míguez

Voy a mostrar la herramienta diagnóstica de ingreso de la Facultad de Ingeniería integrada a los sistemas de información de la facultad, para mostrar no solo el diseño o cómo se utiliza, sino cómo eso está englobado en todos los sistemas y en todas las acciones de la propia facultad, que me pareció que era más interesante. Vamos a ver lo que

nos llevó a diseñar estos sistemas de información, pasamos rápidamente por los distintos instrumentos que tenemos y qué resultados y qué acciones estamos teniendo.

Las consideraciones son obvias, todos estamos embarcados en estos problemas. Es más, no son exclusivos nuestros, son de todo el mundo, hace poco estuvimos en un congreso mundial de educación y vimos que los problemas en las facultades de Ingeniería son los mismos, que tienen problemas en el ingreso, que tienen grandes problemas de abandono, grandes problemas con los muchachos en todo tipo de disposiciones y actitudes más allá de los conocimientos. Al ingreso a la universidad lo que típicamente se señala son limitaciones en estrategias cognitivas en los distintos niveles, distorsionadas por una acción específica, también según los distintos campos disciplinares, y sobre todo hábitos, actitudes y disposiciones hacia el aprendizaje que a veces son dificultosos para trabajar. Esto lleva a un lento avance en las carreras de la facultad que sea, a un bajo índice de egresos y a una elevada tendencia a la deserción, sobre todo en primer año, que es lo que vivimos no solo en Facultad de Ingeniería sino en toda la universidad.

Esto nos para a pensar en lo siguiente: tenemos que poder emplear herramientas diagnósticas que ayuden al sistema a tomar decisiones oportunas y acertadas, nosotros no podemos realizar acciones por las percepciones que tenemos, necesitamos tener más información para poder actuar en consecuencia. En particular esta frase de Clapton: «Si los profesores no saben en qué consiste el aprendizaje y cómo se produce, tienen las mismas posibilidades de favorecerlo que de obstaculizarlo». Nos pareció una premisa interesante para trabajar entre los docentes de Facultad de Ingeniería a partir de la información que se recogía.

En el año 92, la Facultad de Ingeniería ya ponía una prueba diagnóstica de física y de matemática. En el 2001 se empieza a trabajar, se trabajó incluso durante un tiempo con la Unidad de Enseñanza de Facultad de Ciencias en el diseño de una prueba diagnóstica que pudiera ser útil y de la que pudieran hacer uso los docentes de la facultad y también los estudiantes, que no cayera solo en un resultado. Estos sistemas de información que decide implementar la facultad se integran con metodología cualitativa y cuantitativa. En particular la metodología cuantitativa, la herramienta diagnóstica al ingreso, que aplicamos an-

tes de que los estudiantes empiecen los cursos, típicamente el sábado anterior —las clases empiezan el lunes—. Esta herramienta diagnóstica al ingreso fue seguida durante dos años, tuvimos la oportunidad de aplicar una herramienta diagnóstica a los estudiantes con determinado avance en las distintas carreras que ofrece facultad, y eso nos permitía correlacionar y seguir la historia de cada uno. Tenemos un sistema de evaluación docente en la función enseñanza, que también triangula con estos otros instrumentos y nos brinda información de lo que está pasando en las aulas desde un punto de vista cuantitativo.

Si vamos a la metodología cualitativa que nos ayuda a interpretar los números, tenemos: entrevistas en profundidad con los estudiantes a distintas alturas de la carrera, grupos de discusión con los estudiantes y observaciones de aula no participantes. Todo esto se mueve de un lado y otro y permite ir generando un contexto y una adecuación al contexto de lo que pasa con nuestros estudiantes. También seguimos lo que llamamos el avance estudiantil de seguimiento académico, en el que la unidad básica de cada estudiante, de cada individuo, es la generación a la que pertenece. Seguimos cohortes, podemos hacer otros filtros y otras elecciones, pero típicamente lo hacemos por cohortes. El referente es el plan 97, la facultad está creditizada desde esa época, entonces vamos viendo la cantidad de créditos que van obteniendo los estudiantes en el período de tiempo que sea de interés estudiar, y si es necesario también se hacen avances sobre el movimiento académico ya viendo resultados en las asignaturas, en particular.

La herramienta diagnóstica de ingreso se empezó a aplicar en una etapa piloto entre los años 2002 y 2004, porque es muy difícil armar una prueba, todos los que somos docentes acá fuimos estudiantes, sabemos lo difícil que es jerarquizar, seleccionar, decidir. Se hizo entonces una etapa piloto para tener una prueba que fuera confiable y sacar conclusiones a partir de ello que ayudaran a tomar decisiones. En el año 2005 llegamos, después de esta etapa piloto, a una prueba que aplicamos cada año a toda la generación de estudiantes ingresantes, y diseñamos pruebas equivalentes. Tenemos un equipo multidisciplinario que trabaja en la prueba, que está integrado por docentes de matemáticas, de física, de la Unidad de Enseñanza, un equipo de estadísticos, docentes de química. Esta prueba está constituida por múlti-

ples componentes, a las componentes disciplinares más tradicionales, más típicas que la facultad ya tenía, que eran física y matemática, se agrega la componente de química. Hay preguntas abiertas específicamente en matemáticas en las que se ve el desarrollo y se hace toda una categorización de los errores que cometen los estudiantes en esas preguntas abiertas. En física, química y matemática son preguntas de múltiple opción en las que los distractores son seleccionados en la línea que decía Andrés: ese distractor me hace pensar que el estudiante tuvo determinado pensamiento detrás, y me permite deducir o inferir el error que tiene en su manera de resolver el problema.

Tenemos también una componente de comprensión lectora con una parte de múltiple opción muy superficial de reconocimiento en lo que tiene que ver con la habilidad de lengua o de algo que ya está escrito, es muy sencillo. Aun así a algunos estudiantes no les va bien, y esos estudiantes en particular nos preocupan mucho. Después hay una parte de desarrollo, el estudiante tiene que identificar la idea principal de un texto breve y referirla, escribirla con sus propias palabras en un rectángulito que le damos, y eso también se corrige. En esa prueba hay una dificultad mayor, se le exige identificar una idea y expresarla por escrito. Lo que vemos y no es ninguna novedad es que las dificultades en comprensión lectora atraviesan cualquier disciplina y pueden invalidar el aprendizaje de física, de matemáticas, sencillamente porque no entienden lo que leen, y los estudiantes cada vez leen menos.

Después tenemos un cuestionario que aborda perfiles motivacionales y estrategias de aprendizaje en nuestros estudiantes. Porque no solo es importante lo que saben, también lo que pueden hacer, cómo dicen que estudian, qué motivaciones tienen para estar acá, para seguir acá, para estudiar en términos más generales, para trabajar en equipo, entonces poder integrar esas otras dimensiones hace que se enriquezca mucho el conocimiento de cada generación de estudiantes que entran a la facultad. ¿Con qué objetivos se pone esta prueba? Por un lado los docentes de primer año se acercan a sus estudiantes en cada generación. Se acercan a lo que saben, a cómo dicen que estudian, a quiénes son. Tenemos las variables socioeconómicas, pueden tener acceso a esa información. Por otro lado el estudiante también se

posiciona en la facultad, esto es lo mínimo que está esperando y yo lo tengo o no lo tengo. Si no lo tengo, qué hago.

Les voy a contar cómo se usa esa información. Pretendemos favorecer la retención de estudiantes que se quieran quedar, porque hay un abandono que es razonable, «me equivoqué, no me gusta esto, quiero hacer otra cosa». Entonces, se trata de la retención de aquel que sí quiere quedarse allí. A ese queremos ayudarlo.

Pretendemos demarcar tempranamente estudiantes en franco riesgo de fracaso académico y ver qué acciones podemos tener con ellos. Diseñar estrategias de intervención adecuadas al perfil de ingreso y analizar la influencia del ritmo de avance en las carreras.

La herramienta diagnóstica media se aplicó durante dos años. Es también un desafío el diseño de esa herramienta. Se aplica a estudiantes avanzados, a mitad de carrera, definidos con cada una de las carreras. Para los estudiantes que entraron en la generación 2005 y obtienen los créditos, si la herramienta diagnóstica media es en el 2008-2009 puedo comparar, puedo ver su historia. Se vuelven a poner componentes específicas con componentes generales. El cuestionario de motivación de estrategia de aprendizaje entonces me muestra qué está haciendo ahora la institución con los estudiantes que llegaron y cómo llegaron.

Veamos algunos resultados de todo este período y sobre todo las cosas más generales. Por un lado salimos en la prensa cada tanto en algunos titulares, entonces nos llaman y nos dan la oportunidad de hablar de estas cosas y compartir y hacer acciones en este sistema educativo que está tan fragmentado.

Nos permite tener los datos generales de nuestros estudiantes. Nos permite saber de dónde vienen, qué edad promedio tienen, conocer variables socioeconómicas, ver que el 70 o el 80 % de la Facultad de Ingeniería son varones, que vienen la mitad de Montevideo y del interior, que hay un alto porcentaje de estudiantes que vienen de instituciones privadas, algo que es bastante característico en el área de ciencias. Nos permite ver en ese avance los estudiantes que siguen, los de Montevideo, los del interior. Los varones, las mujeres, cómo les va a las mujeres; no les va nada mal.

Ese ritmo de avance, ese perfil de los estudiantes que avanzan también nos permite hacer reflexiones sobre qué es lo que está pasando, y tiene que ver con estos filtros que se mencionaban antes. ¿Hay filtros antes? Sí. ¿Siguen los filtros después? Sí. Y parece que son inevitables mientras no cambiemos algunas cosas.

La herramienta diagnóstica al ingreso nos permite saber cuál es la distribución de puntajes, cómo les va a los muchachos. Nos permite analizar en qué preguntas se equivocaron, por qué se equivocaron, cuáles fueron los distractores. Se hacen el índice de discriminación, el índice de dificultad, el equipo de docentes toma una decisión sobre cómo distribuir esto que comentaba Alicia, ya no entre discriminación y dificultad, sino cómo hago ese interjuego para que no sea tan fácil que todos contesten todo ni tan difícil que nadie conteste nada, en ese interjuego se pone de acuerdo el equipo de docentes que diseñan la prueba.

Esto se conjuga con el avance de los estudiantes, voy viendo año a año dónde están los estudiantes, cuántos créditos obtuvieron, cuáles salvaron los cursos, cuáles no salvaron los cursos, típicamente en Facultad de Ingeniería primer año es muy difícil y a veces están dos tres años para aprobar las asignaturas de ese año.

Hacemos un seguimiento longitudinal de nuestros estudiantes del plan 97, eso lo hacemos para cada generación, y a medida que van pasando los años podemos ir haciendo informes en profundidad de cada generación. Repito que se conjuga la metodología cuantitativa con la cualitativa, porque decir que a un equis porcentaje le va mal en herramienta, que un equis porcentaje perdió un curso no nos dice las causas, puede haber muchas causas, y ahí sí lo cualitativo cobra importancia. Nos permite saber que los estudiantes que trabajan más de 20 horas son los que más desertan, porque la universidad parece no estar adecuada totalmente a las exigencias de los estudiantes que trabajan ocho horas por día. Nos permite ver que la gran mayoría de los estudiantes que tuvieron suficiencia en la herramienta diagnóstica siguen siendo población activa después de cinco años. Podemos saber qué diferencia hay entre estudiantes activos y desertores. Por ejemplo en el puntaje que obtuvieron en HDI o en la cantidad de componentes suficientes, cuantos más componentes suficientes tienen en la prueba, mejor les va en la facultad. Tenemos mucho estudio para adelante,

podemos comparar esos resultados en HDI con lo que va pasando después, podemos hacer relaciones entre el rendimiento y la herramienta y cómo avanza en la carrera la población activa, podemos hacer muchísimos cruzamientos para tratar de entender esto que es muy complejo y en aras de hacer acciones.

El cuestionario de motivación de estrategias de aprendizaje es muy enriquecedor, nos permite ir de la punta del iceberg, que es lo que vemos con los estudiantes y con las disciplinas, a lo más profundo, ver qué pasa con sus cuestiones, con cómo dicen que estudian. Por ejemplo, les preguntamos los motivos y les presentamos una lista cerrada de siete motivos, después de haber hecho toda una parte abierta durante varios años. El motivo que más señalan típicamente, generación tras generación, es el placer que les causa saber más sobre temas que los atraen, y el que sigue es garantizarse inserción laboral, que está muy bien, y en Ingeniería están bastante acertados.

Los muchachos entran con algo que podemos llamar motivación intrínseca en términos superficiales, por un cuestionario otra vez, porque es lo que ellos dicen sobre lo que piensan, lo que dicen sobre ese tema. Y otras preguntas, a modo de ejemplo: «Mientras resuelvo un problema pienso cómo se relaciona con otros problemas de la materia»; muchos dicen que no, porque lo que les interesa es el resultado.

Cómo dicen que aprenden: el 25 % cuando no entiende algo se queda con la duda y no le pregunta a nadie, ni a los compañeros ni a los profesores, ni mira un libro, ni busca en internet, hacen como máximo lo que se les pide y no más, que es un buen ahorro de energía, pero siempre les decimos que lo que ellos entienden que es lo máximo no es lo que el otro entiende que es lo máximo y en esa disonancia a veces tienen problemas.

Casi un tercio estudia de memoria todos los temas, de memoria repetitiva, de memoria no significativa, casi la mitad estudia de memoria lo que no comprende bien. Esto es un promedio de 2005 a 2012.

Un 30 % estudia solo los temas que preguntan siempre. Y el 70 % no estudia con otros ni en grupos, esta es una característica muy fuerte en los estudiantes de Ingeniería, en general son estudiantes muy solitarios. Lo vemos cuando hacemos algunos trabajos en liceos, vemos

que ya empiezan a diferenciarse por esa característica cuando eligen las orientaciones.

Y también nos permite ver la distribución de mitad y final de carrera según las zonas de Montevideo, nos deja ver que están en las zonas donde se agrupan los profesionales en Montevideo. Por eso decía que seguimos reproduciendo el sistema, no estamos logrando salir de eso. Sin embargo con algunas apreciaciones que tomamos en cuenta vamos a tomar algunas acciones. Nos damos cuenta de que el libre acceso no es suficiente para garantizar condiciones de igualdad en el avance académico, es un problema en toda América Latina la permanencia y la posibilidad de egreso sobre todo de aquellos estudiantes de contexto desfavorable, de los cuales a la Facultad de Ingeniería llegan muy poquitos.

Resulta central establecer mecanismos de inclusión que desactiven estos filtros. Ahí encuentro algunos trabajos bien interesantes de nuestro rector, recomiendo uno que habla sobre la divisoria del aprendizaje, que nos parece muy acertado. Dice que la divisoria digital es importante, pero es solo un aspecto de la divisoria del aprendizaje, que es lo que está dividiendo a los países que realmente logran que los muchachos avancen y que la educación vaya para adelante de los que no, y qué se hace con lo que uno aprende, si eso le da posibilidades de cambiar su entorno familiar, de cambiar su entorno social, de pensarlo creativamente. Para hacer estas cosas hay que intervenir deliberadamente en estas situaciones, y eso es lo que está intentando hacer la Facultad de Ingeniería.

Pasamos a algunas acciones en escala micro, pero queremos trabajar con todos los estudiantes que ingresan. Entonces estamos haciendo acciones en escala macro con los recursos que tenemos, de la mejor manera que podemos, pero, como siempre, recursos insuficientes. Sabemos que la entrada a la universidad les está cambiando el ecosistema a los estudiantes, que hay cosas que los complican mucho, sencillamente porque pasan de una institución a otra más grande. Porque venían de un Bachillerato de Ingeniería en el que eran cinco estudiantes y ahora entran y son 1200, que después van a un teórico y son 100 o van a un práctico y son 30, que tampoco es un disparate. Les genera un no lugar y una no pertenencia que hay que trabajar con ellos.

En este sentido queremos utilizar estos sistemas de información, estas pruebas diagnósticas que ponemos como un proceso para mejorar la enseñanza de una manera formativa y que permita los aprendizajes de los estudiantes.

La facultad tiene una actividad introductoria en la que se da la bienvenida a los muchachos, se les explican cosas. Eso nos ha permitido tomar decisiones educativas con información adecuada y oportuna y decir: así como están montados los cursos de primer año, para el 95 % de estos estudiantes que están ingresando no sirven, sirven solo para un 5 %, tenemos que hacer cambios.

Aparecen modalidades lentas de cursados, trayectos diferenciados que se asocian con cursos anuales en los que además de trabajar la disciplina se trabajan estrategias de aprendizaje en matemática, trabajo en equipo. Los muchachos no quieren estudiar en equipo, pero tienen que hacerlo porque después van a trabajar en equipo y se tienen que poder comunicar no solo entre ingenieros sino con cualquiera de la sociedad. También es importante desarrollar esas habilidades, esas competencias, y en eso también estamos trabajando. Trabajamos en la flexibilización física y matemática en ambos semestres, modalidades alternativas. Hay un cursado de física que entusiasma mucho a los estudiantes que se llama Física 1++, que tiene aplicaciones que realmente dan ganas de hacer física de vuelta. Tenemos cursos breves de apoyo en estrategias de aprendizaje, comprensión lectora, matemáticas y lógica en aquellas cuestiones que entendemos que son básicas y con las que los estudiantes no están pudiendo a partir del diagnóstico que estamos haciendo y de las entrevistas.

Desarrollamos un apoyo permanente al estudiante. Damos una devolución de los resultados de la herramienta diagnóstica. No solo se cuelga un papel con el resultado, sino que se les dice «esto dio así, fomentamos que haga esto», esta es la sugerencia de un equipo orientador. En algunos años funcionó, ahora están trabajando los talleres de orientación al inicio, el 1 y el 2, en los que se resuelven un montón de problemáticas de los estudiantes en el sentido de que pueden consultar, preguntar, acercarse, hay un espacio de consultoría y atención permanente que está atendido entre la asistencia académica y la Uni-

dad de Enseñanza, y las tutorías que también son bastante comunes en muchas facultades.

La Unidad de Enseñanza también hace un trabajo con los docentes a partir de estos diagnósticos y del sistema de evaluación docente. Tenemos cursos, tutorías didácticas, para el diseño de las pruebas se trabaja con los docentes, porque no es tan sencillo hacer una buena evaluación de múltiple opción, es fácil de corregir pero es muy difícil de diseñar, y se están generando materiales alternativos y libros para los cursos de inicio.

En resumen, la integración de las distintas dimensiones que implica la estrategia que aplica la facultad con las pruebas diagnósticas integradas a ese sistema nos resulta un importante factor explicativo, formativo y de mejora continua para diseñar estrategias institucionales adecuadas al contexto para cada generación que ingresa, para poder promover una mejor enseñanza y un mejor aprendizaje. Estos son dos aspectos:

Primero, las herramientas cubren algunos aspectos de la realidad, es difícil cubrir todo, ponemos muchos componentes, muchos sistemas, muchos instrumentos, pero es difícil cubrir todo. Por otra parte, los sistemas son limitados y hay que seguir tratando de que evolucionen en el diseño, en los alcances, en la aplicación, en los recursos que tenemos, con las acciones que hacemos, porque quizás hay competencias en las generaciones nuevas que no estamos pudiendo medir porque todavía no nos dimos cuenta. Estos estudiantes *multitask* que están haciendo una cosa y uno piensa que están distraídos a lo mejor no están tan distraídos, algunos de ustedes están contestando un mensaje en el celular y me están escuchando, entonces nuestros estudiantes pueden hacer lo mismo. Nos está faltando todavía capacidad de medir eso, y creo que lo tenemos que tratar de desarrollar.

Debate

Pregunta 1

—Con respecto a la exposición de Peri, ¿quiénes y cómo elaboran las pruebas? ¿Se hace una evaluación de los aspectos creativos de los niños?

Pregunta 2

—Para Marina: tú describiste muy bien todo lo del aparataje, todo lo que es la estrategia, pero de los resultados sale que eso no frena la deserción; ha mejorado el nivel académico de los estudiantes.

Pregunta 3

—¿Las pruebas son en papel?

—Sí.

Pregunta 4

—Con respecto a los resultados que todos conocemos de Ingeniería, que siempre son cuestiones más puntuales de lo que presentó la unidad, es infinitamente más rico lo que sale con relación a lo que se hace con todo esto que es infinitamente más global de lo que se discute normalmente, todos hemos estado preocupados estos años por los resultados y seguimos. Las pruebas diagnósticas de Ingeniería han ido descendiendo, fue el planteamiento central de la presentación de Marina. Conocemos los resultados, que en los últimos años han llegado a niveles francamente alarmantes en el sentido de que algo pasa. Tenemos insuficiencia a nivel 1 y entonces, ¿qué reflexión hacen? Genuinemente todos nos podemos preguntar si no hay algún elemento que tenga que ver con la formulación de los instrumentos de esa evaluación, con el instrumento, con el corte del instrumento.

Otro problema que me preocupa desde hace mucho es el contexto en el que se aplican las pruebas. Esta prueba y cualquier otra en la universidad, y lo mismo va a ser para la comprensión del año que viene. Las pruebas de conocimiento se aplican al inicio sin situación de examen, los muchachos no se preparan para eso, están acostumbrados a ser evaluados continuamente con una preparación previa. To-

dos sabemos cómo inciden las formas de preparación de las pautas en los resultados. Entonces el hecho de que se apliquen al inicio, sin una adrenalina y una recuperación de conocimientos para ese fin, parece que puede estar incidiendo en los resultados.

Pregunta 5

—En cuanto a lo que decía Marina, que desertan porque «esto no es lo mío», ¿cuántos se han distanciado —si lo han visto— por esa primera mala experiencia? ¿Cuánto está influenciado por ese primer test que les dice «esto no es lo tuyo»? ¿Cuánto de eso se ha tenido en cuenta, cuánto de este costo-beneficio, cuánto de este beneficio no tiene un alto costo? Habría que cuidar las dos caras de la moneda en el sentido de qué le pasa al estudiante cuando ve que su resultado es cero, ese es el costo.

Por otro lado, cuando hablás de los equipos, cuando conformamos el equipo, qué perfil tiene ese equipo, quién es la gente que arma la prueba, si está en primero o no está en primero, si son los mismos que después les dicen a los estudiantes «no sigan porque no vale la pena, no van a poder con esto». No sé si no hay que tener cuidado con la conformación de esos equipos y esas pruebas, sumado a lo que Mercedes expuso como llamado de atención en lo que es la prueba de ingreso, que comparto. Son cuestiones que hay que mirar.

Pregunta 6

—Si no entendí mal, las dos experiencias tienen una composición fundamental que está soportada por múltiple opción. Lo que planteó Andrés de las temáticas de los distintos decanos, el extremo, las cuestiones que tienen que saber, las cuestiones de la asignatura, las cuestiones propias de la asignatura, hasta desempeños, en competencias, de conocimiento. Mi duda es si eso es parte de lo que había hecho la universidad unos años atrás, gran parte de lo que importa son esas habilidades para leer, para aprender, para desempeñarse en ambientes diferentes, etcétera. No tenía que ver directamente con saber tanto de matemática, tanto de física, etcétera. Mi pregunta es si se está evaluando eso, o se están evaluando los desempeños que tenemos inten-

ciones de conocer y esas capacidades de los estudiantes no se están evaluando.

Pregunta 7

—Apoyo las dos preguntas anteriores. Ustedes dieron una cantidad de datos de suficiencia e insuficiencia, pero quedó poco claro qué pasa en el aula, qué les dijeron los profesores en cuanto ustedes supieron los números. ¿Eso refleja lo que el docente debe saber de aula?

Pregunta 8

—Ampliamos y hacemos la descripción de la estrategia: si a nivel del cuerpo docente eso generó algún tipo de reflexión, algún tipo de estrategia, si sirvió para algo y qué resultados arrojó.

Pregunta 9

—Dos aspectos que vienen de la mano de los citados. Una cosa que me pregunto con respecto a la confiabilidad de los números, simplemente pensando en los números, es la motivación o el interés del estudiante que lo hace contestar bien, ese esfuerzo que pone el estudiante en la prueba diagnóstica para resolver cualquier situación planteada. No es lo mismo errar que no contestar, al menos en el servicio donde estoy hay muchos ítems que no se contestan y a veces nos queda esa interrogante: ¿por qué no se contestan?, ¿porque estoy apurado, porque estoy muy cansado y no tengo ganas de pensar o porque no lo sé? Eso me parece que fue un hilo de sombra.

El segundo aspecto, que es el más preocupante, al menos en el servicio donde trabajo, es qué se hace con el número o qué se hace con el contenido que induce el número, con la información, que puede ser con una doble mirada como planteaba Alicia. Me parece que lo interesante es que esas miradas se incorporen al menos en los primeros ciclos, es decir: por qué se equivocó y cuáles podrían haber sido los motivos distractores que planteaba Marina que llevan a tener un error, cuáles son los tipos de errores y qué puedo hacer con esos errores en mi trabajo en aula cuando inician la universidad. Me parece que ese es el desafío más grande en la enseñanza.

—Una pregunta muy general para Alicia, un pedido de comentarios sobre lo que escuchó de los dos compañeros, dada su especialización, ¿cómo ve estas experiencias?, al menos por lo que pudo escuchar aquí, y en particular, ¿cómo ve la construcción de consensos, como decía Andrés, que implica la construcción misma de las pruebas, si es que eso se visualiza? Y a ellos, que nos expliquen cómo se construyen las pruebas en este sentido, no tanto en el aspecto técnico sino los acuerdos que implican sobre lo que se supone pertinente aprender.

Una segunda cuestión, respecto a esta maravilla del pasaje a lo digital, si no hay un riesgo de limitar algunas posibilidades. Potencia mucho, eso quedó muy claro, pero ¿no implica también algunos límites, especialmente a lo cualitativo, a lo que no es múltiple opción, a lo que queda abierto? Vos mostraste al final las 1500 respuestas abiertas de los docentes, pero es bastante difícil procesar eso, ¿qué se hace con los estudiantes y sus respuestas abiertas, si es que las hay en este sistema? Es decir, si esta modalidad no termina condicionando hacia cierto tipo de pruebas y deja otras de lado, si tiene algún límite y, en todo caso, si ustedes están buscando cómo superarlo y qué han hecho en ese sentido.

Para Marina, aunque podría ser para cualquiera: parecería que las pruebas intentan medir conocimientos y en cierta medida después aprendizajes. ¿Cómo se relaciona eso con la enseñanza? Tú lo sugeriste cuando dijiste «se han cambiado cursos a partir de allí». ¿Qué tanto evalúan de la enseñanza, qué tanto evalúan de la actividad docente y cómo están impactando en la actividad docente? Se preguntó por allí por impactos de las pruebas en ese sentido.

Y finalmente, para todos, me parece que buena parte de la preocupación por las pruebas diagnósticas en los últimos años y por las pruebas en general tiene que ver con una percepción muy preocupada y preocupante de los resultados generales de los sistemas educativos. Como ya preguntó algún compañero: ¿estamos logrando cambiar algo a partir de esto, o más bien terminamos o bien en una gran depresión o bien, como dijo Andrés, en un traslado de culpas hacia otros, todo viene mal de algún otro lado, con estos estudiantes tan desastrosos nada podemos hacer? Ese tipo de diagnóstico catastrófico puede ser real, pero no nos dice mucho sobre qué hacer y cómo.

Respuestas

Alicia Camilloni

Creo que todas las evaluaciones, que muchas veces tienen la característica de tener que ser prácticas, es decir de poder evaluar con cierta calidad y rapidez para obtener la información, surgen mucho de la ganancia, entonces uno tiene que saber que siempre son parciales, que no se va a poder evaluar todo y por lo tanto que no conoce acabadamente ese objeto que se está evaluando.

En el caso de Andrés, me parece claro que se evalúa una gran cantidad de conocimiento de los estudiantes, con lo cual se da información que no existía. Me pareció interesante que hablara de alineamiento —respecto de qué se está evaluando— de los conocimientos de los alumnos con los currículos y también con las estrategias de enseñanza. No me cabe duda de que es absolutamente ventajoso, mi primer comentario sería ese, poder encarar eso y pensarlo más a lo largo del tiempo como un proceso con el cual se van mejorando las pruebas. Ciertamente, a mí también me preocupa qué se dice y qué se pone, cómo son las pruebas, no creo que con lo largo aparezca lo técnico. Y por otro lado lo pertinente, creo que el problema de la evaluación es resolver técnicamente bien la evaluación pertinente que uno tiene que hacer respecto a los contenidos. En ese sentido, los felicito por el avance y particularmente porque uno sabe que no es todo, que siempre hay más para evaluar.

Respecto de la presentación de Marina, me quedé con alguna pregunta. Por ejemplo, me gustaría saber si los alumnos conocen los contenidos que van a ser evaluados en la prueba diagnóstica, y me preocupaba más una pregunta que también se hizo: ¿los alumnos se preparan para esto?, ¿cómo se preparan, cómo llegan a la prueba? En buena medida las respuestas van a depender de esto, sumado a las expectativas previas que puedan tener. En segundo lugar, los contenidos que se toman, ¿realmente pertenecen al currículo escolar? Porque lo único que sé acerca de los resultados es lo que vi en algunos titulares, por ahora no tengo más información, son bastante impresionantes. Estamos acostumbrados a ese tipo de titulares en los diarios. Se hizo

una investigación precisamente sobre el carácter catastrófico que se les da a esos titulares en los diarios, con distintos objetivos.

Si solamente ese porcentaje de alumnos ganaron la aprobación... Voy a hacer una pregunta que quizás parezca técnica. Vos hiciste una referencia a los índices de dificultades y a los índices de discriminación, ¿cómo hacen ustedes para fijar los índices de dificultad y los índices de discriminación? Esa es una de las cuestiones fundamentales. ¿Hay un proceso de estandarización previo de la prueba diagnóstica? Me parece muy interesante la trayectoria que van haciendo de los objetivos que se proponen con la prueba, pero siempre se corre el riesgo de que los resultados sean contradictorios, que no se logren algunos efectos, que se produzcan otros. Yo había dicho que a mí sí me preocupan las pruebas diagnósticas por el fracaso inicial del alumno. El sentimiento de fracaso como sentimiento que revive situaciones personales, y en segundo lugar como una especie de señal que le da la institución de «esto no es para vos», que es lo que finalmente algunos de los estudiantes entienden. Es un caso muy interesante el de ingeniería, personalmente me interesa mucho, qué pasa con las carreras de ingeniería.

En el mundo faltan ingenieros, hay programas muy importantes de promoción para la incorporación de nuevos estudiantes a la carrera de ingeniería, para lo cual se hace un trabajo muy importante en algunos países tratando de convencer a las mujeres de que la matemática puede ser aprendida por la mujer. Hay un prejuicio en otros lugares, creo que no pasa ni en Uruguay ni en Argentina, pero hay programas destinados específicamente a convencer a las mujeres de que pueden aprender matemática y pueden seguir ingeniería, porque ciertamente uno observa cómo, a pesar de que las mujeres se han incorporado masivamente a la universidad, esto no ha ocurrido con algunas carreras de ingeniería. Ustedes tienen el 70 % de varones y tendríamos que ver qué pasa con eso. El mensaje «ingeniería es difícil, ingeniería no es para todos» es un mensaje duro, que quizás vaya en contra de lo que ustedes se proponen y todos nos proponemos, que es tener una mayor matrícula en la carrera de ingeniería, abrir las puertas a los estudiantes, de manera que se puedan incorporar y tengan confianza en que pueden hacerlo, es decir, contribuir a un autoconcepto positivo

respecto de sus posibilidades y un mensaje que es contradictorio de autoeficacia negativa que tenga que ver con la versión inicial.

Aunque ustedes no le den el carné de calificación, aunque no sea una prueba vinculante respecto a la decisión que tome, el estudiante, con toda su historia escolar, arrastra varias evaluaciones. El peso que las evaluaciones tienen como señales, como respuestas que dan las instituciones respecto de lo que ellos saben —esto está estudiado en el análisis del discurso del docente, el docente siempre evalúa lo que el alumno contesta—. El fracaso, el no poder contestar o el contestar mal en pruebas iniciales, puede ser muy duro. Esto es lo que les quiero preguntar, si han evaluado cuáles son los sentimientos o las modificaciones que puede haber en el estudiante que llega con una expectativa frente a éxitos o fracasos en la prueba diagnóstica.

Andrés Peri

La primera pregunta era quién hace las pruebas. Tenemos itemólogos, que son maestros que hacen ítems de evaluación. Hacemos un llamado público a maestros que quieran participar, los maestros participan en una capacitación de cuatro días en producción de actividades de evaluación y después entran en un proceso en el cual, a través de la plataforma que desarrollamos, envían propuestas de evaluación y un equipo técnico de la unidad dice «está bien la consigna», «está mal la competencia», «es ambiguo lo que quiere medir, vuelva a hacer», interaccúa hasta que al final el equipo de especialistas aprueba el ítem y les pagamos. El proceso de producción tiene algunos líos, no es perfecto, pero tiene algunos mecanismos de seleccionar a los buenos en la producción descentralizada, y después en la selección de esas actividades para conformar la prueba. Pero después nosotros tenemos que asegurarnos de que las actividades que tenemos cubren toda la tabla y cubren los distintos niveles de dificultad. Eso es difícilísimo, esos son líos que enfrentamos.

La segunda cuestión son dudas sobre la múltiple opción. Hay una investigación que en su momento presentó Pedro Ravela titulada *Qué evalúan los docentes en primaria cuando ellos proponen actividades*. Hay un discurso en los docentes, cuando dicen «las evaluaciones están dadas, son todas memorísticas de baja producción», hay que ver las activida-

des que hacen. Cuando se analizaron las actividades que proponen, son memorísticas, repetitivas, de contenidos específicos de las cosas que han enseñado. En este debate entre cuestiones de múltiple opción o respuesta construida —ojo, últimamente hay una pregunta sobre las limitaciones—, es cierto que nosotros no les decimos a los docentes «no haga pruebas de múltiple opción en su aula porque está mal». Porque un docente con 22 alumnos contesta cada una de las producciones y ve la producción. Cuando uno hace una prueba de sistema, no de aula, es otro partido, yo a veces digo que hay como una ruptura epistemológica de la producción, de lo que hay que hacer. Es distinta la prueba de aula, la que yo hago para ver progreso y cómo lo que yo enseñé de esta señal que construye este tipo de pruebas. La primera consigna de la evaluación de aula es que uno tiene que evaluar aquello que enseñó, ergo, no puede haber ninguna evaluación por fuera del proceso de enseñanza. Eso es para la evaluación de aula, para cualificar alumnos en un grupo.

Pero los sistemas educativos complejos son algo más que la suma de aulas, el aprendizaje institucional de un sistema educativo no funciona sumando aulas y sumando propuestas de evaluación de docentes, ni sumando notas, ni sumando promociones. Funcionan por otros y eso es lo que me parece más importante tratar de transmitir, y es difícil porque la mayoría de los docentes se forman en el IPA con ese paradigma, y entender el lugar que ocupa esta evaluación, cómo nos ayuda para otras cosas es un bicho raro y molesto en esa cuestión y me parece uno de los desafíos más importantes que tenemos.

Después, sobre el impacto en los docentes, dos cosas. La primera, cuando uno hace cuestiones a esta escala es bravo evaluar el impacto. Cuando hicimos la primera evaluación los docentes nos decían «¡qué lindo esto!, pero estoy en mi casa robándoles el tiempo a mis hijos analizando los resultados». Es muy claro que tiene que haber tiempo institucional para pensar qué pasó. En la segunda versión fuimos al Consejo de Enseñanza Primaria y explicamos que no podíamos seguir desarrollando este proyecto sin tiempo institucional de reflexión en las escuelas. Secundaria tiene ya salas de reflexión instituidas en todos los centros, pero en Primaria depende de la categoría de escuela, las escuelas de tiempo completo tienen toda la semana, las de tiempo

extendido tienen dos veces a la semana, las escuelas comunes y las de práctica no tienen tiempo institucional para reflexionar sobre el proceso educativo. Eso es un problema enorme, no de las evaluaciones, sino de las mejoras, y parte de lo que pasó es que el consejo resolvió cuatro salas de reflexión a posteriori del ciclo, dos para la evaluación en línea y dos para los aspectos que no evalúa la evaluación en línea.

Yo digo que la magia pasa en cada escuela. La responsabilidad nuestra es construir un recurso significativo, potente. Alicia plantea cómo lo alinea, para mí el verbo alinear es fundamental, pero es difícilísimo en nuestro sistema. ¿Están alineadas las pruebas con la didáctica de los formadores en servicio? No lo puedo asegurar. ¿Están alineadas las pruebas con el programa? Sí, lo puedo asegurar. Pero el programa no llega a determinar exigencias, ese es un problema. ¿Cómo mejoramos? Yo soy de la tradición de hacer hipótesis de trabajo, hacer, pensar, reflexionar y seguir. Si hacemos una educación teórica sobre el mundo y sus alrededores, la actividad de los triángulos es la que esperamos para cuarto, es riquísima la discusión sobre eso. Recuerdo las discusiones con la inspección técnica, la primera prueba de segundo que hicimos en su momento, había una cuestión de que las capas de comprensión lectora frente a un texto son muy variadas, lo que uno quiere en la prueba diagnóstica es ver niveles de desempeño, no niveles de suficiencia. Los niveles de suficiencia son para otro tipo de pruebas, pero esa es una discusión en esta cuestión del carácter de las pruebas. Todas nuestras pruebas están colgadas en internet, si alguien quiere verlas, va al sitio y están.

Yo antes tenía una ilusión marciana, quería que la prensa discutiera las actividades que proponíamos y hacer públicas esas actividades, pero la prensa nos dio como adentro de un gorro el año pasado. Y ojo, cuando dijeron «la mitad de los escolares fallaron en la evaluación en línea» fue porque salimos a aclarar que si es una prueba formativa no hay que fallar. La prensa después dijo «ahora que les fue mal dicen que son formativas». Nunca podés ganarle a la prensa. Otra cosa que nos pasaba cuando hacíamos las evaluaciones era que como después es un proceso continuo lo pueden hacer en cualquier momento, como publicábamos las respuestas, los niños eran más astutos, iban a los lugares y contestaban basándose en la información que ya estaba publicada.

Entonces pasamos a publicar la prueba pero no los resultados, antes publicábamos los resultados.

Ahí hay una cuestión sobre los impactos. Creo que el impacto es modesto en este sentido. Nuestro objetivo es primero que el desarrollo profesional de los maestros y los docentes sea una actividad significativa dentro de un año lectivo normal, hago una actividad y me junto con otros para pensar qué pasó. Para mí ese es el desarrollo profesional, no es sacar al maestro del aula y volver a enseñarle matemática, ese es el modelo tradicional que hemos ensayado siempre. Para mí lo crítico es alinear, pero después están el Prolee y el Prorrazona, dos programas transversales de cuño e influencia de la universidad en la ANEP. ¿Están alineados con nosotros? No. ¿Queremos alinearlos? Sí. Pero ¿cuánta gente comprende la estrategia de intervención?, porque implícitamente en esta cuestión la evaluación es una forma de intervenir. Reconozco que no solo se evalúa, detrás hay un formato de conceptualización, comprensión, posicionamiento y reflexión.

Recuerdo que cuando presentamos las evaluaciones Héctor Florit hizo dos preguntas. Dijo: «Andrés, ¿lo que preguntamos en la prueba es lo fundamental que tiene que saber un alumno en el grado?». Yo le dije: «Hicimos muchos procesos de acuerdo y de construcción de consenso para la tabla de especificaciones, uno puede ir al sitio y ver la tabla de especificaciones, que es la matriz a partir de la cual se arma la prueba. Hicimos los procesos de acuerdo, eran públicos antes de la prueba». Y la segunda pregunta fue: «¿Y ese es el nivel de exigencia socialmente esperado para el grado?». Le dije que no podía contestarle en una respuesta. Las actividades eran hipótesis de trabajo, cuestiones relevantes, pero después cuando analizamos la prueba, el nivel de exigencia era muy alto. Ojo, ¿qué tituló la prensa? «Una forma de mejorar los resultados educativos: hacer pruebas más fáciles». Les preguntamos a más de 1500 docentes y en matemática y ciencias más del 40 % consideró la prueba muy difícil o difícil.

Después estaba la prueba del límite de lo digital, que cierra cosas y abre otras puertitas, porque uno puede tener una animación como reactivo a lo que quiere hacer, puede tener una simulación como desencadenante de la propuesta. Nuestras pruebas no tenían colores, eran en blanco y negro, uno podía arrastrar un objeto y hacer una función

biunívoca en algo de relacionar algo con otro, se abría un horizonte de otras posibilidades. Lo cierto es que creo que las pruebas que hacemos todavía son tontas, son pruebas que se podrían hacer en papel, todavía no tenemos desarrollo de estas capacidades. La posibilidad está. Pero quiero transmitir lo difícil del proyecto global, comprender que lo estamos haciendo en un sistema público y privado.

Segundo, construir pruebas es difícilísimo, que estén alineadas, que sean significativas y que tengan el nivel de exigencia socialmente esperado.

Tercero, tener una plataforma en la que todo el que quiera hacer la prueba la pueda hacer. Eso supone pelearme con Ceibal, decirle «esta escuela no tiene conectividad, tiene que tenerla, porque tiene que hacer la evaluación en línea». Y después los problemas en las escuelas donde hay muchas máquinas rotas y el esfuerzo que supone que para esta actividad todos los niños tengan sus computadoras cargadas con las baterías para poder hacerla en la clase.

Unir todas esas cuestiones es parte del desafío de este tipo de proyecto. Ahora el desafío es construir el trayecto de tercero a tercero. Para mí es muy importante llegar a acuerdos conceptuales. El problema es dar el salto entre primaria y secundaria, una parte es el cambio institucional, como pasa también en este salto entre secundaria y la universidad. Pero estas cuestiones son conceptualizaciones, progresiones, profesionales distintos esperando cosas distintas, los mismos muchachos. Yo tengo en la presentación una gráfica que muestra la repetición de primero, segundo, tercero, cuarto, quinto, sexto de primaria, primero, segundo y tercero de secundaria. Muestra que en el período 2002-2011 en primaria bajó sistemáticamente todos esos años, y en secundaria subió. Cuando vean la gráfica se van a dar cuenta del impacto que tiene el problema de las transiciones.

Marina Míguez

Es vital dónde pone uno el foco. Como decía Andrés recién, la prensa pone el foco donde le interesa para vender un diario, tenemos ese sufrimiento compartido. Por eso mismo hacemos informes que van mucho más allá de ese titular, que están colgados en la página de la Facultad de Ingeniería y se han distribuido ampliamente, en los que

el foco no es «el 92 % es insuficiente» como titula la prensa. Hacemos análisis no solo de las componentes disciplinares; las preguntas se suelen focalizar en el nivel de suficiencia en matemáticas, en física y en química, y en la Facultad de Ingeniería, si bien esas componentes están, tenemos el foco puesto en un cuestionario de motivación de estrategia de aprendizaje. Tanto que hay una tesis de maestría y una tesis de doctorado atrás de ese cuestionario, y ahí puso las pilas la facultad que tiene que ver con los sentimientos. Claro que nos interesan los estudiantes como individuos y por eso aparecen todas esas situaciones, la prueba permite tener una información para saber qué acciones desarrollar que sean consonantes. Sino, podríamos hacerlo por el olfato, «me parece que ahora tendríamos que hacer no sé qué o trabajo en equipo». Pero ¿es lo que precisan los muchachos?

Otra pregunta era sobre el contexto en el que se aplica la prueba diagnóstica. Voy a citar a una docente de la Facultad de Ingeniería: «Lo mejor de la prueba diagnóstica es el contexto en que se aplica, porque así aprenden lo que es dar un parcial en la Facultad de Ingeniería con algo que no tiene que ver con la asignatura», con una prueba que no implica perder una asignatura. Aprenden que tienen que traer la cédula de identidad, por ejemplo —año tras año cada vez más estudiantes vienen sin cédula—, y con los padres. Esos son cambios que se ven y esta docente, a la que le interesan mucho los temas de enseñanza, dice que para ella eso es realmente algo valioso.

El contexto también es importante porque se les explica lo que quiere decir esta herramienta diagnóstica. En la actividad introductoria a los muchachos se les dice «no tienen por qué estudiar para la prueba». No queremos que estudien porque no es una prueba que evalúa conocimiento, es una prueba que intenta medir, mediando las vacaciones, de qué se acuerdan y cómo vienen, qué traen puesto a la facultad, ese es el objetivo de la prueba. Quizás es una prueba diagnóstica atípica, quizás no es la prueba diagnóstica que quiere medir el nivel de suficiencia que nos interesa, que se calcula por un método, cada equipo docente lo calcula y después hay una suficiencia global de la prueba. ¿De qué te acordás después de las vacaciones, qué traés puesto, qué conocimiento traés puesto? ¿Te acordás de sumar, por ejemplo, media-

namente? En vacaciones todos nos acordamos de sumar, ¿te acordás de sumar fracciones? No, de eso ya no se acuerda.

Tú preguntabas sobre los contenidos. Los contenidos son de los programas que se dan en cuarto, quinto y sexto de bachillerato. El equipo que diseña la prueba se integra con docentes de matemática y de física de primer año de facultad y de quinto y sexto de bachillerato, docentes de química, también de bachillerato y de facultad, inspectores de Secundaria de química, de física, de lengua, es decir de los distintos componentes de la prueba. Nosotros nos reunimos con los inspectores todos los años, ven las preguntas y dicen si les parecen adecuadas o no, hacen sugerencias, «acá convendría cambiar el modo en que se pregunta, porque los estudiantes no están acostumbrados a verlo de esta manera y sí de otra». El equipo está integrado por esa gente y docentes de la Unidad de Enseñanza, psicólogos, especialistas en informática, en estadística, etcétera.

El cuestionario de motivación de estrategias de aprendizaje está diseñado por psicólogos de la unidad, es algo en lo que ponemos mucho énfasis porque nos permite conocer las particularidades de los estudiantes de Facultad de Ingeniería, que son muy particulares frente a estudiantes de otras orientaciones, tienen un locus de control interno bastante fuerte que les permite posicionarse frente a las actividades de facultad pensando que todo lo que hagan va a ser importante. Tienen las metas definidas, tienen una autoconfianza particular y que nosotros conocemos, se les aplica el alcance de la herramienta y la utilidad de la herramienta. Me atrevo a decir que la relación costo-beneficio es muy buena, que el beneficio es alto. A los estudiantes les va mal en la prueba y no les importa. No nos vienen a decir «me siento un fracasado» o «esto no es lo mío». Eso no se escucha en el típico estudiante de ingeniería, porque nosotros los atendemos en los talleres de orientación al inicio y hablamos con ellos. Están en el espacio de orientación y consulta, hablan con los asistentes académicos y no surgen estas cosas, surgen otras. Al contrario, el estudiante se siente más contenido en facultad por las acciones que se vienen desarrollando. Algunas empezaron este año, o el año pasado. Tuvo que haber un desarrollo histórico para poder comprender el fenómeno, que es muy complicado, y además hacer lo que podemos hacer con los recursos que tenemos.

También preguntaban si se había mejorado el nivel académico. La duración de las carreras está disminuyendo. No me atrevo a decir por qué, porque es un fenómeno multicausal, pero hemos logrado bajar un promedio de trece años a once años y ahora estamos en ocho años y medio. Todo esto tiene que ver, puede ser, pero son muchas cosas que se van sumando. Los cursos también han ido cambiando, acompañando este perfil del estudiante que ingresa. El docente en el aula se encuentra con estudiantes que no saben algunos contenidos, pero si saben estudiar, «¿esto es lo que no viste?, tomá, estudialo acá», y ya está. Pero si no saben estudiar o tienen un problema de estrategias de aprendizaje, tenemos que hacer otro tipo de apoyos y en eso es que la facultad está tratando de poner las pilas.

Preguntaban algo sobre la formulación de la prueba. Ya les conté cómo se define el punto de corte, cada equipo docente señala esto, pero sobre todo después con los resultados de la prueba trata de comprender en qué cosas se equivocaron los muchachos y por qué, esto de los distractores que yo decía que hay que seleccionarlos de determinada manera. No tenemos un porcentaje muy alto de no respuesta, los muchachos contestan, tienen tres horas para todo y se quedan las tres horas.

Preguntaban sobre la motivación para responder las pruebas. Cuando empezamos con esto la prueba diagnóstica era nada, era voy y con testo. Recuerdo desde el primer año que estuve en la Facultad de Ingeniería, a la media hora los estudiantes se levantaban y entregaban la prueba diagnóstica que se ponía, con problemas de matemáticas y de física. Y recuerdo que un estudiante me dijo: «Tome, profesora, ¿sabe qué? Hace calor, juega Peñarol, me quiero ir». Entonces dijimos: este es un subdiagnóstico, obviamente. Le dimos muchas vueltas, y finalmente apelamos a una motivación intrínseca que creemos que está funcionando y que a todos nos mueve: aquellos estudiantes que llegan al nivel de suficiencia global obtienen puntos para acumular en los parciales en las asignaturas de primer año, del primer semestre. La motivación funciona porque a todos nos funciona, y combinada con la buena motivación intrínseca que tienen los muchachos pareció ser una buena forma, porque están las tres horas tratando de contestar, responden todo y en el cuestionario son confiables las respuestas que dan, que después se convalidan con otros instrumentos y con las entregas.

Tercera parte

¿De qué evaluación formativa hablamos? Presentación y debates sobre experiencias de evaluación formativa en la Universidad de la República

Este encuentro se realizó el
29 de mayo de 2014 de 9 a 12 h y 14 a 18 h
en el Salón de conferencias de la Facultad de Ciencias Económicas
y de Administración
Gonzalo Ramírez 1926 - Montevideo
Organizado por:
cse - Unidad Académica
Red UAE
UAE de la Facultad de Humanidades y Ciencias de la Educación

Conferencia: Mag. Rebeca Anijovich
Universidad de Buenos Aires, Universidad de San Andrés y
Facultad Latinoamericana de Ciencias Sociales, Argentina

Debates pedagógicos sobre evaluación formativa en el nivel universitario

Rebeca Anijovich

Para comenzar nuestro recorrido me interesa reconocer la relación que hay entre evaluación, la enseñanza y el aprendizaje. Más específicamente quiero referirme a la necesidad de articulación, de búsqueda de coherencia entre cómo se enseña, cómo se aprende y cómo se evalúa. Muchas veces observamos ejemplos de buenas prácticas de ense-

ñanza que no se condicen con las prácticas de evaluación. Por ejemplo, enseñar utilizando casos pero, a la hora de evaluar, formulamos cinco preguntas para que los estudiantes respondan lo que plantean los textos en forma de evocación de información.

Por eso la primera afirmación que comparto con ustedes es la necesidad de lograr que nuestras prácticas de enseñanza y de evaluación sostengan el mismo enfoque teórico, de tal modo que se favorezca y no se obstaculice el aprendizaje de los estudiantes.

Dentro del vasto campo de la evaluación nos referiremos específicamente a la evaluación formativa, un término que propuso Scriven en 1967 para diferenciarlo de la evaluación sumativa. Para este autor, la función de la evaluación formativa estaba en comprender los procesos de aprendizaje de los estudiantes para rediseñar o ajustar las estrategias de enseñanza.

Son muchas las investigaciones que hoy encontramos acerca de la evaluación formativa. Linda Allal (1980); Royce Sadler, (1998); Alicia Camilloni (1999, 2004,); Edith Litwin (1999; 2008), Black y Wiliam (1998, 2004); Gordon Stobart (1997); Lorraine Shepard (2000); Menucha Birnbaum, (2003); Lorna Earl (2003); Susan Brookhart, (2003); Ricky Stiggins, (2001); Lucie Mottier López (2006, 2008, 2011); y Philippe Perrenoud (2008), entre otros, han iluminado aspectos diversos del campo de la evaluación formativa. Esta herramienta, en tanto, procura contribuir a la mejora de los aprendizajes de los estudiantes y al aumento de la probabilidad de que todos aprendan. Consideramos que la evaluación formativa debe servirle al estudiante para comprender mejor de qué modo se comporta en el rol de aprendiz, qué habilidades tiene, qué le resulta fácil y difícil, y qué recorridos hace para aprender algo en profundidad. Al mismo tiempo debería servir para monitorear la buena marcha de la enseñanza reuniendo la información que permita modificar lo necesario para favorecer los aprendizajes de los estudiantes.

Compartimos algunas ideas significativas que caracterizan a la evaluación como formativa:

- Los profesores comunican con claridad a sus estudiantes los propósitos y los objetivos de la materia, y establecen relaciones

claras con las tareas académicas y los criterios de calidad de las mismas.

- Los profesores ofrecen y/o construyen con los estudiantes criterios explícitos de evaluación.
- Los profesores ofrecen retroalimentaciones variadas en sus estrategias y frecuentes en el tiempo, focalizando en aquello que los estudiantes pueden hacer en sus futuras tareas.
- Los estudiantes se ofrecen retroalimentaciones entre sí.
- Los profesores y estudiantes son conscientes del impacto emocional de las retroalimentaciones en la autoestima y en la motivación.
- Los profesores estimulan y promueven en los estudiantes procesos metacognitivos y reflexiones sobre sus trabajos. La intención es que sean los alumnos quienes asuman un trabajo activo de monitoreo y comprensión de sus propios procesos de aprendizaje, estrategias, obstáculos, avances; es decir, que se apropien de sus aprendizajes de tal modo que no le atribuyan al azar o al profesor sus logros o sus dificultades. La investigadora francesa Linda Allal (1980) ha sostenido que la evaluación formativa tiene una función reguladora que permite identificar las debilidades y las fortalezas del aprendizaje de un estudiante. Esta función «transita entre dos polos. El primero es externo y recae en el docente, quien a partir del análisis y la reflexión de las actividades de los estudiantes ofrece explicaciones alternativas, ejemplos, sugerencias. El otro polo es el de la regulación interna de los propios estudiantes, que se pone en evidencia cuando realizan actividades de aprendizaje y de reflexión sobre las mismas». (Anijovich y González: 2010: 11)
- Los profesores ajustan su enseñanza a partir de las informaciones que recogen de sus observaciones, del análisis de las producciones y aportes de sus estudiantes.

William (2009), investigador del campo de la evaluación, ha reformulado sus primeras definiciones al plantear que «la práctica en una clase es formativa en la medida que la evidencia acerca de los logros de los estudiantes es obtenida, interpretada y usada por docentes, aprendices o sus pares para tomar decisiones acerca de sus próximos

pasos en la instrucción que tengan probabilidades de ser mejores, o mejor fundadas, que las decisiones que ellos hubieran tomado en la ausencia de la evidencia que fue obtenida». La contribución más importante que aporta esta definición está vinculada a las evidencias que se buscan obtener. Por ese motivo es necesario decidir cuáles serán las evidencias que permitirán demostrar lo que los estudiantes lograron y que, al mismo tiempo, serán útiles para continuar con la enseñanza; si no se tuvieran esas evidencias, no se podrían efectuar ajustes, mejoras y avances. Entonces una pregunta que nos podemos formular es cuántas evidencias necesito recoger de los estudiantes y de qué tipo. David Allen (2000), investigador de Harvard, se formula una pregunta interesante: «¿Qué harán los profesores con todos los papeles que juntan de sus estudiantes a lo largo de un año?». En tal sentido, plantea que reunimos más de lo que podemos procesar. La intención, entonces, es revisar y ser criterioso al momento de decidir qué informaciones hay que recoger.

Al respecto, sumamos los aportes de Klenowsk (2009), quien plantea que la evaluación formativa tiene que ver con el diálogo, los intercambios, las demostraciones y las observaciones realizadas para mejorar los aprendizajes, y propone el uso de un portafolios³ como un asistente de la evaluación.

Es importante destacar, además, que la evaluación formativa cumple también una función clave al identificar la brecha entre lo que habría que lograr y lo logrado, y así encarar un trabajo de mejora y revisión para disminuir esa grieta.

Si los procesos de feedback son para mejorar, entonces es necesario ir más allá de una visión de transmisión a una de construcción; de un modelo de recepción pasiva a un modo activo de ofrecer y recibir, intercambiar y encontrar sentido al contenido del feedback. En definitiva, de preguntar y preguntarse.

El poder de la retroalimentación no reside sólo en cuándo y cómo se ofrece, sino también en cuándo y cómo se recibe (Andrade 2010; Hattie and Gan 2011; King, Schrodts, and Weisel 2009).

3 *Colección de trabajos con el propósito de evidenciar los aprendizajes logrados, los procesos y recorridos realizados.*

La autoevaluación y la evaluación entre pares

Pensamos en un estudiante autónomo cuando dispone de competencias para enfrentar desafíos, resolver problemas e interactuar con otros. Son, en este sentido, valiosos los aportes que la autoevaluación y la evaluación entre pares pueden ofrecer al desarrollo de las capacidades de un alumno autónomo.

Para que un estudiante pueda autoevaluarse requiere, entre otras cuestiones, una retroalimentación que lo ayude a clarificar los objetivos que debe alcanzar, como también los criterios y estándares en su formación profesional. Es importante que esta retroalimentación no sea únicamente oral sino que se ofrezcan documentos escritos que permitan al estudiante utilizarlos como referencias.

Planteamos tres preguntas a modo de instrumento que nos pueden ayudar en esta tarea:

1. ¿Adónde voy? La pregunta remite a los estudiantes a que piensen acerca de «para qué voy a hacer esto, cuál es el objetivo de este trabajo, cuál es el sentido de resolver esta consigna, qué se supone que debería aprender haciendo esto, qué estrategias voy a usar, qué recursos, qué tiempo.
2. ¿Cómo lo estoy haciendo? Los enfrenta a pensar acerca de qué tienen que ajustar, qué ayuda necesitan, qué tendrían que cambiar.
3. ¿Cómo continúo? La idea del desafío es mirar hacia el futuro.

Cuando trabajamos con propuestas de autoevaluación, debemos primero analizar con los estudiantes cuál es el beneficio de encarar este proceso. También es necesario en este punto confrontar creencias y concepciones erróneas, discutir estrategias diferentes para hacerlo, mostrar ejemplos y convertir esta práctica en una actividad rutinaria.

Con relación a la evaluación entre pares (Topping, 2009), esta herramienta muestra el valor de la evaluación en el nivel universitario, que culmina en la metacognición a través de la cual los estudiantes hacen explícitas sus estrategias de aprendizaje y autoestima en relación con la creencia y la confianza en su capacidad para aprender.

La evaluación entre pares recoge muchas críticas en diferentes estudios, entre las cuales podemos reconocer: 1) que los aportes son

superficiales, focalizados en lo que gusta o no gusta de los argumentos; 2) que están poco o nada relacionados con lo pedido para la tarea; 3) que tiene mucho peso el vínculo establecido entre quien ofrece y quien recibe retroalimentación; 4) que la dimensión afectiva prevalece sobre los aspectos referidos a la tarea y el contenido.

Teniendo en consideración las dificultades mencionadas, se postulan algunas prácticas como:

- Sugerir que las retroalimentaciones entre pares sean anónimas,
- Que se trabaje sobre un protocolo o una guía de preguntas,
- Alternar retroalimentaciones por escrito con devoluciones cara a cara,
- Organizar devoluciones de un equipo a otro.

La calidad de la evaluación

Podemos afirmar que la calidad de la evaluación depende de varias cuestiones, entre ellas, de la calidad de las consignas de trabajo y de la calidad del feedback.

En primer lugar, nos referiremos a la calidad de las consignas. Abogamos por objetivos de trabajo desafiantes, vinculados con el contexto y el campo profesional, que les permitan a los estudiantes utilizar el conocimiento de manera flexible, estableciendo relaciones entre diferentes áreas disciplinares.

Una consigna, entonces, deber ser auténtica. Es decir, demandar una tarea que tenga sentido para el sujeto que la realizará y le permita aprender en una situación lo más cercana posible al «grado óptimo de incertidumbre», aquel nivel que lo desafía pero no lo paraliza. Cuando los estudiantes encuentran sentido en las propuestas de los docentes aprenden más y mejor y se comportan de manera más autónoma.

Pensar las consignas como oportunidades para que los estudiantes expliquen, muestren y den cuenta ellos de lo que están comprendiendo en lugar de evocar información que otros han dicho.

En segundo lugar nos referimos al valor del feedback en la evaluación formativa. Entendemos el feedback como un proceso de diálogo, de intercambio, de modelar, de formulación de buenas preguntas.

Pensemos en diversidad de estrategias para ofrecer feedback. No se trata de «decirle algo al estudiante para que lo haga». Esta manera de entender el feedback está alineada con todas las pedagogías del diálogo entre profesores y estudiantes, del buen clima en el aula y la institución.

Podemos agrupar, a modo de presentación sintética, tres enfoques diferentes sobre el feedback:

- 1) Las que se plantean acerca de cómo modifica los procesos de pensamiento y los comportamientos. Si los estudiantes, a partir del trabajo con el feedback, comienzan a pensar distinto sobre la disciplina y sobre los modos de aprender (Black y Wiliam: 2009; Shute: 2008; Macfarlane: 2006; Ecclestone: 2005; Perrenoud: 1998).
- 2) Las que muestran cómo la información reduce la discrepancia entre lo que se entiende y lo que se pretendía ser comprendido (Sadler: 1989; Hattie y Timperley: 2007).
- 3) Las que enfocan en la motivación como un factor significativo de los aprendizajes. El feedback tiene un impacto emocional por lo tanto se tiene que ser cuidadoso en qué y cómo y en cuándo y dónde se ofrece y bajo qué marco teórico. (Shute: 2008).

Nos interesa señalar dos distinciones acerca del foco de las retroalimentaciones. Aquellas enfocadas a la autoestima, como por ejemplo «eres un buen estudiante» o destacar el desempeño y el esfuerzo que buscan influir en los aspectos emocionales de los estudiantes. Este tipo de retroalimentaciones puede provocar mejoras en el aprendizaje del estudiante, a partir de un incremento en la autoconfianza y capacidad de logro. También puede inducir un efecto contrario, apoyado en comentarios positivos, el estudiante no reconozca que necesita mejorar algunos aspectos de su trabajo.

Una pregunta que nos abre la retroalimentación refiere al tema del tiempo. ¿La retroalimentación se propone mirar hacia el pasado o trabajar sobre el futuro? ¿El impacto y el cambio posterior a la retroalimentación perduran? Es importante tener presente que el estudiante no siempre va a producir una acción espontánea de revisión y/o mejora por el solo hecho de haber recibido una retroalimentación. Esto puede deberse a que muchas veces el estudiante no comprende qué proble-

mas presenta su trabajo y, en este sentido, las orientaciones o sugerencias que se le puedan ofrecer no resultarán apropiadas para mejorar sus aprendizajes. O simplemente que el estudiante «acepte» la retroalimentación sólo para cumplir con lo que solicita el profesor. Para que se puedan revisar resultados y/o desempeños hacia atrás y al mismo tiempo tender a una mejora en el futuro, es necesario dar continuidad al proceso de retroalimentación. Esto es posible, por ejemplo, solicitando al estudiante que vuelva a entregar su trabajo con las modificaciones indicadas o que diseñe un plan de mejora para su próximo trabajo.

Una última pregunta que se desprende de lo planteado antes se refiere a las decisiones que toma un estudiante con respecto a la información que recibe: ¿La comprende? ¿La considera para efectuar alguna acción? ¿La ignora? Parece existir la suposición de que alcanza con ofrecer retroalimentación para que los estudiantes comprendan y emprendan acciones a partir de ella. Pero investigaciones como las de Higgins (2001) muestran las dificultades de los estudiantes para comprender las retroalimentaciones y de los docentes para ofrecerlas en un lenguaje claro y accesible. La efectividad de una retroalimentación depende del tipo de aceptación de quien la recibe, pero también de cuán eficaz sea quien la comunica. El que emite el feedback debe utilizar un nivel de lenguaje verbal y no verbal adecuado para su receptor, y crear un contexto físico y emocional adecuado para que el mensaje impacte en la dirección deseada.

La evaluación formativa y la reflexión de los docentes

La evaluación formativa nos ofrece, en tanto, información sobre la enseñanza, por eso nos parece una oportunidad interesante para favorecer la reflexión de los docentes sobre sus propias prácticas. Esta posibilidad implica la generación de espacios y tiempos para que los docentes conformen una comunidad de aprendizaje que reflexione sobre sus prácticas. En soledad es muy difícil lograrlo.

La práctica reflexiva en el oficio de enseñar conlleva una gran capacidad para reflexionar durante y después de la acción. No se trata de una reflexión episódica, sino de una postura del profesional docente: crear los espacios, tiempos y dispositivos para el análisis de la práctica.

Bibliografía

- Allal, L. (1980). «Estrategias de evaluación formativa: concepciones psicopedagógicas y modalidades de aplicación». *Infancia y aprendizaje* n.º 11, Madrid.
- Allen, D. (2000). *La evaluación del aprendizaje de los estudiantes* Editorial Paidós. Buenos Aires.
- Andrade, H. (2010). Students as the definitive source of formative assessment. In *Handbook of formative assessment*, ed. H. Andrade and G. Cizek, 90-105. New York, NY: Routledge.
- Anijovich, R. y González, C. (2010). *Evaluar para aprender*. Buenos Aires. Aique.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education*, 5(1), 7-74.
- Black, P. et al. (2004). *Assessment for Learning. Putting it into practice* Open University Press, Glasgow.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment, in Segers, Dochy, Cascallar, eds. *Optimizing New Methods of Assessment: In search of Qualities and Standards*. Dordrecht, The Netherlands: Kluwer, pp. 13-36.
- Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22 (4), 5-12.
- Camilloni, A. (2004). Sobre la evaluación formativa de los aprendizajes. *Quehacer Educativo*, Montevideo (Uruguay) Año XIV n.º 68 (pp. 6-12).
- Earl, L. (2003). *Assessment as learning: using classroom assessment to maximize student learning*. Thousands Oaks. Corwin Press.
- Ecclestone, K. (2005). *Understanding assessment and qualifications in post-compulsory education and training: Principles, politics and practice*. 2nd ed. Leicester: NIACE.
- Hattie, J. y H. Timperley (2007). The power of feedback, *Review of Educational Research*. 77, n.º 1, p. 31.
- Hattie, J., and M. Gan (2011). Instruction based on feedback. In *Handbook of research on learning and instruction*, ed. R.E. Mayer and P.A. Alexander, 249-71. New York, NY: Routledge.
- Higgins, R., Hartley, P. & Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback, *Teaching in Higher Education*, 6(2), 269-274.
- King, P.E., P. Schrodt, and J.J. Weisel (2009). The instructional feedback orientation scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education* 58, no. 2: 235-61.
- Klenowski, V. (2005). *Desarrollo de portafolios para el apje y la evaluación*. Madrid. Narcea.
- Litwin, E. (2008). *El Oficio de Enseñar. Condiciones y Contextos*. Buenos Aires, Paidós.
- Moro, Stella M. (2005). *Formular, leer, analizar, comprender consignas escolares*, Rosario. Universidad Nacional de Rosario.
-

-
- Mottier Lopez, L. (2006). L'évaluation située des apprentissages : le rôle fondamental de la médiation sociale. In G. Figari & L. Mottier Lopez (Ed.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 82-89). Paris: L'Harmattan.
- (2011). Evaluación formativa de los aprendizajes de los estudiantes: síntesis crítica de los trabajos francófonos en Anijovich (comp). *La evaluación significativa*. Buenos Aires. Paidós.
- & Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation*, 3, 5-34.
- Perrenoud P. (2008). *La evaluación de los estudiantes. De la producción de la excelencia a la regulación de los aprendizajes*. Entre dos lógicas. Colihue. Buenos Aires. Pp126-127.
- Sadler, R. (1998). Formative assessment: revisiting the territory, *Assessment in Education: Principles, Policy and Practice*, 5 (1), 77-85.
- Sadler, D. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, 18, 145-165.
- Scriven, M. (1967). «The methodology of evaluation», en Tyler, R. Gagne, R & Scriven, M (eds.), *Perspectives on curriculum evaluation*, Chicago, Rand McNally, pp. 39-83.
- Shepard, L. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*. Vol. 29 n.º 7 October, (1-14).
- Shute, V. J. (2008). Focus on formative feedback *Review of Educational Research*, 78(1), 153-189.
- Stiggins, R. (2007). Assessment through the student´s eyes. *Educational Leadership*, 67 (8), 22-26.
- Stobart, G y Gipps, C. (1997). *Assessment: A teacher´s guide to the issues*. Londres. Hodder and Stoughton.
- Taras, M. (2002). Using assessment for learning and learning from assessment, *Assessment & Evaluation in Higher Education*, 27 (6), pp. 501-510.
- Topping, K. (2005). Trends in peer learning. In K.Wheldall (Ed), *Developments in educational psychology: How far have we come in 25 years?* (pp 59-73). London. Routledge. Falmer.
- (2009). Peer assessment. *Theory into Practice*, 48 (1), 20-27.
- William, D. (2009). Una síntesis integradora de la investigación e implicancias para una nueva teoría de la evaluación formativa. En *Archivos de Ciencias de la Educación*. Año 3-nº 3-4 Epoca-2009. La Plata. Universidad de La Plata.
-

Resúmenes de las Presentaciones

Evaluación formativa en un curso de Histología y Biología del desarrollo mediante plataforma EVA en Facultad de Veterinaria-Udelar

Graciela Pedrana, Claudia Borlido, Fernanda Alcaide, María Helen Viotti, Victoria Riaño, Patricia Genovese, Paula Lombide, Juan Calvo, Elisa Souza, José Passarini, Alejandro Bielli.

Resumen

En el presente trabajo se describe una experiencia de evaluación formativa por portafolio mediante la plataforma EVA (Entorno Virtual de Aprendizaje) en el curso 2013 de Histología y Biología del Desarrollo, en el primer año de la carrera de Doctor en Ciencias Veterinarias de la Universidad de la República. Se generó un «portafolio virtual» conteniendo evidencias de aprendizaje, retroalimentación de docentes, formulaciones y reformulaciones de trabajos por parte de los estudiantes.

La propuesta fue optativa y no sumativa sobre la nota final del curso. Se planteó la propuesta de evaluación formativa durante las actividades prácticas presenciales. Se explicó las tareas a realizar dentro de la plataforma EVA con una consigna definida de subir archivos que fueran evidencias de aprendizaje. La evaluación formativa por portafolio involucró una valoración cualitativa e individual, en la que participaron el 20% de los estudiantes y un total de 113 tareas realizadas en forma continua durante 11 clases prácticas. Se remarcó a los estudiantes el carácter formativo con la finalidad de poder mejorar las metodologías y los recursos didácticos. La evaluación fue contextualizada, con la incorporación de elementos del entorno socio-cultural, tales como los teléfonos celulares y cámaras digitales utilizados para capturar imágenes de los preparados histológicos observados durante las clases prácticas. A los estudiantes, no se les dio pautas específicas de formato para los trabajos del portafolio, siendo libre la elección del tipo de archivo, solo teniendo el límite de capacidad pre-establecida por la plataforma EVA. Luego de finalizada cada actividad práctica, los

estudiantes realizaron la subida de archivos y los docentes realizaron la retroalimentación de cada trabajo. Las devoluciones involucraron los criterios establecidos en una rúbrica generada por los docentes.

Se realizaron sugerencias y correcciones a los trabajos presentados por los estudiantes, pudiendo reformular su trabajo en base a las recomendaciones de los docentes. Las evidencias de aprendizaje en los portafolios variaron desde imágenes capturadas con celular de esquemas de pizarrón, dibujos de cuaderno y textos adjuntos numerando las estructuras histológicas observadas, imágenes capturadas con celular a partir de los oculares de los microscopios durante la clase, solas o incorporando texto adjunto, referencias e indicadores (fechas, círculos, líneas evidenciando las estructuras) hasta textos explicativos de los preparados observados a los distintos aumentos.

En conclusión, 1) este trabajo permitió a los docentes valorar el aprendizaje a través de evidencias de desempeño durante el proceso de enseñanza-aprendizaje, 2) las evidencias de aprendizaje en las actividades prácticas del curso de Histología y Biología del Desarrollo permitieron la retroalimentación del docente, formar juicios, comparar pautas y criterios utilizados para tomar decisiones sobre evaluación y seguimiento de los procesos de enseñanza-aprendizaje durante las actividades prácticas, 3) se promovió el aprendizaje por medio del desarrollo de habilidades como síntesis y selección de material representativo de sus evidencias de aprendizaje, 4) se estimuló el desarrollo de metacognición a lo largo del curso. Se espera que este tipo de evaluación formativa por portafolio se incorpore como una nueva herramienta en otros cursos para permitir tomar decisiones pertinentes durante el proceso de enseñanza-aprendizaje.

Evaluación Formativa en un teórico de Física. Facultad de Ingeniería-Udelar

A. Auyuanet, L. Amy, N. Barrios

Resumen

Física 1 es la primera asignatura de Física que deben cursar los estudiantes al ingresar a la Facultad de Ingeniería. Dicha asignatura tiene una carga horaria de 6 horas semanales: 2 clases teóricas de 1:30 horas de duración cada una y 2 clases de resolución de ejercicios de idéntica duración. Las clases teóricas son masivas, con una concurrencia promedio de 150 estudiantes; tradicionalmente, dichas clases son exclusivamente expositivas.

Los niveles de aprobación de la asignatura son bajos y también existe un importante nivel de deserción en el primer semestre. Intentando revertir esta situación se decidió implementar en un teórico de Física 1 técnicas de Aprendizaje Activo de manera de motivar y comprometer a los estudiantes en el proceso de enseñanza-aprendizaje.

Dentro de las técnicas empleadas, tiene un papel importante la Evaluación Formativa, ya que nos permite obtener una retroalimentación por parte de los estudiantes. Según Fink⁴ una buena retroalimentación debe cumplir con los siguientes requisitos:

- Frecuente: Brindar la retroalimentación diaria, semanalmente, o con la máxima frecuencia que le sea posible.
- Inmediata: Obtener la retroalimentación de parte de sus alumnos tan pronto como le sea posible.
- Discriminatoria: Aclarar bien las diferencias entre trabajos deficientes, aceptables y excepcionales.
- Leal: Mantener una empatía mientras se brinda la retroalimentación a los estudiantes.

Para realizar dicho cometido hemos implementado dos tipos de prácticas dentro de las actividades del teórico: 1) incorporamos la resolución de ejercicios en la clase teórica por parte de los estudiantes; de esa manera rápidamente el docente detecta si el tópico que se ha

4 Fink, D. (2008). *Un Guía Auto-Dirigida al Diseño de Cursos para el Aprendizaje Significativo*. Recuperado de: <http://www.deefinkandassociates.com/Spanish_SelfDirectedGuide.pdf>.

explicado fue comprendido y con qué profundidad. A su vez, los estudiantes toman conciencia de lo que entendieron y sobre todo, de lo que no entendieron. Esto permite en el momento, hacer ajustes o volver a explicar desde otro punto de vista los conceptos teóricos que fueron trabajados previamente. 2) Exhibimos videos con situaciones relacionadas a los conceptos teóricos explicados, y se les pide a los estudiantes que analicen y expliquen la situación mostrada. Los estudiantes a medida que analizan el video discuten sobre los conceptos físicos que deben ser tenidos en cuenta para la explicación del fenómeno; en dichas discusiones van descartando las hipótesis equivocadas y toman conocimiento de sus niveles de comprensión del tema.

Estos dos procedimientos han sido bien acogidos por los alumnos y permiten un desarrollo del teórico más centrado en las necesidades y posibilidades de los estudiantes, que las que brinda el formato tradicional de clase teórica.

Bibliografía

Fink, D. (2008). *Un Guía Auto-Dirigida al Diseño de Cursos para el Aprendizaje Significativo*. Recuperado de: <http://www.deefinkandassociates.com/Spanish_SelfDirected-Guide.pdf>.

Experiencia de Evaluación Formativa en la Facultad de Ciencias Sociales

Laura Paulo Bevilacqua, Inés Martínez, Dorelí Pérez

Resumen

Esta propuesta busca re-organizar la modalidad de interacción entre docentes y estudiantes, para crear un proceso educativo basado en el trabajo colaborativo y el análisis de conocimiento en una comunidad mixta de aprendizaje (grado y posgrado).

Para ello, se continuará con la adecuación de estrategias didácticas para el tratamiento de los contenidos del curso, cuestión que se viene realizando desde 2011. Las dos principales innovaciones propuestas son: a) generalización de la evaluación procesual mixta que se aplicó en parte del curso, el pasado año y b) trabajo conjunto entre estudiantes de grado y estudiantes de posgrado en fase de especialización.

En forma sintética, las líneas de trabajo están dadas por el desarrollo de los módulos temáticos y la puesta en práctica de la antes referida evaluación procesual mixta. Es procesual porque hace transitar a los/las estudiantes por diferentes etapas del análisis de las temáticas y es mixta pues combina participación libre y obligatoria, trabajo virtual y presencial, constitución aleatoria de los grupos y auto-constitución, intercambio oral y producción escrita.

Mediante estas innovaciones, se espera abrir espacios en los cuales los/as estudiantes puedan realizar planteos que respondan a sus necesidades así como a consignas planteadas por el equipo docente. Se espera además, que las actividades oficien como guía para identificar aspectos sustantivos y que faciliten la organización del estudio. El intercambio en grupos al azar, tiene el propósito de que las contribuciones individuales, favorezcan el contacto con otras formas de expresión y de pensamiento, el procesamiento de posturas que dinamicen los aprendizajes. Por su vez, con respecto a instancias en grupos auto-constituidos, se espera que favorezcan la cohesión, aumenten la cooperación, generen mayor productividad y por tanto, otro tipo de situación de aprendizaje. Por último, de los trabajos individuales, se espera textos académicos que condensen los conocimientos procesados en las otras modalidades propuestas en el curso.

En definitiva, pensar en la combinatoria de lo individual y lo colectivo, lo presencial y lo virtual, lo elegido y lo asignado, tiene el propósito de generar entornos diversos que den oportunidad a las fortalezas de cada estudiante a la vez que equilibren la incidencia de sus debilidades (horarios, reflexión inmediata, expresión oral, comprensión y producción de texto académico). Consecuentemente, tiene el propósito de promover la consideración de diferentes perspectivas para reflexionar sobre los temas y sobre los propios aprendizajes.

La articulación con el posgrado, procura institucionalizar esta modalidad de trabajo, para aprovechar la producción interna de la Unidad Académica, vincular el aprendizaje en aula con el análisis de recientes producciones especializadas, reconocer la producción de los autores dándoles valor didáctico como herramienta que fomenta un aprendizaje activo.

La evaluación formativa: una práctica compleja. INET-CFE

Alejandro Miños Fayard

Resumen

Los cursos de Didáctica de las carreras docentes se divide en dos ejes: el curso teórico de Didáctica, y el curso práctica docente. Durante la práctica docente, se realizan visitas de orientación didáctica a los estudiantes. En estas instancias se produce un intercambio entre el profesor de Didáctica, el profesor adscriptor y el estudiante, reflexionando sobre el desarrollo de la clase y los supuestos teóricos de la misma.

El curso tiene un fin formativo, procurando que el practicante mejore su desempeño docente, al tiempo que permiten al profesor obtener insumos para mejorar el proceso educativo. Sin embargo, y más allá de la intencionalidad de los actores, existen algunos problemas (en el sentido de situaciones a considerar y resolver) que deben ser tenidas en cuenta a la hora de proceder a este tipo de evaluación:

1. Grupos de práctica reducidos, donde es posible el tratamiento personalizado. En efecto, las instancias de visitas y el trabajo personalizado con los estudiantes, es un proceso que solo es posible en tanto y cuanto se disponga del tiempo suficiente para ello.
2. Profesores que reúnan los requisitos para desempeñarse como profesor de Didáctica. La evaluación formativa es por definición una instancia de aprendizaje, no solo para el alumno, sino también para el docente, el cual debe poseer la formación necesaria para conocer y aplicar un proceso de este tipo.
3. Entender y compartir, como profesor, el proceso de evaluación como formativo. Partiendo del supuesto que este tipo de evaluación procura la personalización del proceso de enseñanza-aprendizaje (dentro de las posibilidades académicas y normativas), el profesor deberá compartir el hecho que las visitas son más que una sucesión de instancias aisladas de calificación, de la cual solo se obtienen insumos para análisis estadísticos. Despojarnos de la concepción dominante que la evaluación es nor-

mativa, calificativa y clasificadora, está asociada a un proceso intransferible e interno de cada individuo, en particular de los profesores.

4. La evaluación formativa necesita no solo de profesores comprometidos con ella, sino que también requiere estudiantes dispuestos a analizar críticamente su proceso de aprendizaje, dispuestos a autoevaluarse. Desde esta concepción se hace difícil diferenciar un estudiante que no ha alcanzado los mínimos requeridos por falta de trabajo teórico-práctico, de aquellos que no se han mostrado comprometidos con el proceso.
5. Por último mencionamos el hecho que la evaluación está atada a tiempos académicos, los cuales no siempre coinciden con los del individuo. Es así que deberíamos aceptar el hecho que son relativos los niveles de suficiencia académicos, en algunas áreas del conocimiento, pues estos se encuentran limitados por aspectos normativos.

En síntesis, la práctica de la evaluación formativa presentan desafíos importantes, los cuales no solo requieren un tiempo de aprendizaje tanto para profesores como para alumnos, sino que además es necesario diferenciar evaluar formativamente de no evaluar.

Bibliografía

Didáctica, currículo y evaluación. J. Álvarez.

Didáctica práctica para enseñanza media y superior. E. Fiore, J. Leymonié.

Más didáctica. J. Steiman

Un sistema de evaluación con múltiples componentes para sostener el aprendizaje.

Cátedra de Matemática – Facultad de Arquitectura

Victoria Albarenque, Santiago Ballester, Pierino Birriel, Inés Bouvier, Santiago Donadini, Omar Echenique, Omar Gil, Federico Giménez, Gastón Ibarburu, Ana Laura Nuín, Marcel Perchman, Emiliano Prieto, Mary Reimondo, Gustavo Ríos, Luis Rosa, Ramón Sellanes

Resumen

La Cátedra de Matemática de la Facultad de Arquitectura tiene a su cargo el curso de Matemática, de 11 créditos, en el primer semestre de la carrera. Típicamente atiende una población de unos 600 estudiantes en los primeros semestres de cada año, y de unos 400 los segundos semestres.

Este colectivo se organiza en grupos de aproximadamente 100 estudiantes, que a su vez se dividen en equipos de trabajo de 6 estudiantes. En el aula suelen trabajar entre dos y cuatro docentes (número que frecuentemente incluye uno o dos colaboradores honorarios), buscando aplicar estrategias de enseñanza activas, basadas en el reconocimiento de que el aprendizaje resulta del quehacer del estudiante en interacción con sus pares.

En cada dictado de cursos se propone un sistema de evaluación que combina instancias individuales y grupales, pequeñas pruebas en clase, parciales con posibilidades de recuperación que requieren la revisión en equipo del trabajo de cada estudiante miembro del equipo, tareas grupales y oportunidades de autoevaluación y evaluación por pares. En esta presentación expondremos el sistema de evaluación que estamos usando durante el primer semestre de 2014 para la acreditación del curso.

También discutiremos algunos antecedentes, posibles líneas de modificaciones y profundizaciones futuras y otras componentes de evaluación no sumativas a las que hemos recurrido. Discutiremos brevemente una experiencia reciente de curso de verano, en el que esta estrategia de evaluación se combinó exitosamente con la atención a crear y mantener un buen clima de aula.

Son referentes directos para el diseño de nuestros sistemas de evaluación la organización del aula según el modelo de Aprendizaje Basado en Equipos (*Team Based Learning*), desarrollado por Larry Michaelsen, y los métodos de Instrucción por Pares propuestos por Eric Mazur y su grupo.

Evaluación formativa en cursos de bases biológicas de odontología. Unidad de Apoyo a la Enseñanza.
Facultad de Odontología

Mariana Seoane

Resumen

La carrera de Odontología desde el año 1994 implementa un modelo curricular inspirado, entre otros aspectos, en enfoques de enseñanza activa y evaluación formativa. La Unidad de Apoyo a la Enseñanza presenta resultados globales de su implementación en el primer año de la carrera para la cohorte 2009 a nivel del curso básico biológico (Morfofunción). Esta área de formación es la que presenta los mayores niveles de rezago en la etapa inicial de la carrera.

A partir de datos obtenidos de la tesis de Maestría en Enseñanza Universitaria sobre el desempeño académico de esta cohorte en los dos primeros años de formación, se realiza un análisis del vínculo entre la evaluación continua (EC) y los rendimientos de los estudiantes en términos de aprobación final del curso.

La cohorte 2009 mostró en el pasaje de primero a segundo año un nivel de rezago considerable, llegando únicamente el 30% a segundo año. La estrategia de cursado prevaleció en el área de Morfofunción en casi la mitad de los rezagados activos, hecho que se constata de forma usual en los cursos de primer año de la carrera. Dentro de los factores que pueden incidir se encuentra la posibilidad de obtener una calificación de cursado (EC) que beneficie al estudiante en la instancia de la evaluación final (EF).

La evaluación formativa se encuentra enmarcada en el principio de enseñanza activa y contempla una valoración permanente del es-

tudiante durante su cursado. Presenta un componente objetivo (implementación de parciales, pruebas) y uno subjetivo (valoración del comportamiento del estudiante en clase, participación, compañerismo, asiduidad, compromiso, atención al paciente, etc.). Actualmente no es clara la asociación que existe entre la *EC* y la *EF*. Apreciaciones generales sugieren que la calificación que el docente brinda al estudiante podría estar fuertemente incidida por el componente subjetivo.

Esta evaluación se manifiesta en una calificación de *EC* al final del cursado de una Unidad de Aprendizaje (UDA) según la escala 0 a 12 (DDD a SSS). Su incidencia en el rendimiento académico (*EF*) varía según la modalidad de aprobación de curso prevista en el Plan de Estudios (examen, exoneración). En el área de Morfofunción incide en la aprobación de la *EF* cuando su valor es de 8 o más.

En este primer avance interpretativo se presenta cómo se comporta la *EC* en esta área de formación, según la normativa curricular vigente, considerando el contexto de cursado y características propias de la disciplina.

Se observa una asociación significativa entre los resultados de la *EC* y la *EF* para las tres UDA a lo largo de los tres primeros periodos de examen, a mayor valor de *EC* más oportunidades de aprobar la *EF*. La posibilidad de obtener una *EC* de 8 o más es beneficiosa para los estudiantes ya que permite la aprobación de la *EF* con un porcentaje menor al requerido con una calificación *EC* de 7 o menos. Alrededor de un 30% de estudiantes con calificación *EC* de 8 a 10 se vieron beneficiados por su continua y lograron así aprobar la *EF*, mientras que ningún caso de *EF* de 11 o 12 mostró dicho efecto (no necesitó de la calificación *EC* para aprobar).

Es posible considerar que el proceso realizado por el estudiante a lo largo del cursado es congruente con su calificación de *EC* y con el rendimiento a nivel de la *EF*, identificándose un pequeño porcentaje de casos en situación de necesitar el beneficio de la *EC* para la aprobación final del curso; todo lo cual motiva considerar otras variables incidentes en el proceso. La dinámica de evaluación en el área de Morfofunción puede afectar la decisión de recurrar de los estudiantes que se rezagan en el primer año con el fin de obtener una *EC* que los beneficie en el resultado final del curso.

Experiencia sobre Evaluación Formativa en Pediatría. Facultad de Medicina

Mónica Sosa

Resumen

En este caso la descripción de una experiencia sobre evaluación formativa en el aula Universitaria se desprende de la representación del sistema didáctico. El currículo y la evaluación atraviesan todos los integrantes del aula y se evidencia la complejidad de una situación de enseñanza-aprendizaje en particular.

Se desarrolló en el anfiteatro del Centro Hospitalario del Norte Gustavo Saint Bois en el área de la Facultad de Medicina de la Universidad de la República con 9 estudiantes del tercer año del Ciclo Clínico Básico Comunitario (CBCC) en el área de Pediatría del plan de estudios del año 2009 de la Facultad de Medicina.⁵

El objeto de conocimiento fue la aplicación de la estrategia para la Atención Integrada a las Enfermedades Prevalentes de la Infancia (AIEPI) de la OPS/OMS en niños de 2 meses a 4 años de edad⁶ en la consulta ambulatorio.

Al inició se analizó el material gráfico en cartelera y su aplicación en la consulta, se examinó el respaldo bibliográfico y audiovisual de la OPS/OMS - UNICEF apoyados en casos clínicos.

Desde el inicio se aplicaron los instrumentos de evaluación con el objetivo de generar un aprendizaje significativo; se utilizó una observación guiada por una hoja de cotejo pre codificada (anexo1) frente a cada caso clínico y un cuestionario (anexo2) con siete preguntas objetivas y tres que favorecieron la reflexión sobre lo aprendido, vivido y sentido en el aula. Los instrumentos de evaluación son válidos, confiables, prácticos y útiles.

5 *Plan de Estudios 2009. Facultad de Medicina.*

6 *Manual guía. AIEP/OPS/OMS.UNISEF.AÑO 2010. Disponible en: <http://www.dem.fmed.edu.uy/materno/Pediatrica/index_aiepi.htm>.*

Bibliografía

- Bordas, M; Cabrera, F. (2001). Estrategias de evaluación de los aprendizajes centrados en el proceso. *Revista Española de pedagogía*. n° 218. Enero-abril. pp. 25 a 48.
- Camilloni, Alicia R. W.; Celman, Susana; Maté, Carmen y Litwin, Edith (1998). *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires. Paidós.
- Camilloni, Alicia (2004). «Sobre la Evaluación Formativa de los aprendizajes» *Quehacer Educativo*, Montevideo. Año XIV n.° 68, pp. 6-12.
- House, Ernest R. (1994). *Evaluación, ética y poder*. Madrid. Morata.
- (1992) «Tendencias de la evaluación» *Revista de Educación*. n.° 299. Madrid.

Formatos de evaluación del Proyecto Integral «Protección Social, Instituciones y Práctica Profesional». Licenciatura en Trabajo Social.

Elizabeth Ortega, Celmira Bentura, Alicia Tommasino, Alejandro Mariatti, Cecilia Sapriza, Maria Jose Beltran

Resumen

Presentaremos las diferentes formas de evaluación que conforman el Proyecto Integral, «Protección Social, Instituciones y Práctica Profesional».

El Proyecto Integral, es una asignatura teórico práctica que se desarrolla durante los 4 últimos semestres de la licenciatura en Trabajo Social, que articula la formación de la intervención profesional y de la investigación.

Participamos 6 profesores que trabajamos con los mismos criterios de evaluación.

La evaluación consta de 5 productos en el primer año y 6 en el segundo, algunos individuales y otros colectivos.

Algunos de estos productos son parciales tradicionales en aula donde se evalúan los contenidos trabajados en el teórico, otros son realizados en domicilio en base a pautas guías que procuran organizar la sistematización y reflexión teórica de la intervención pre profesional. Estas pautas son diferentes de acuerdo al momento del proceso de aprendizaje y de intervención. Algunos de estos son de elaboración individual y otros colectivos. Por último deben realizar un proyecto de investigación y un informe de los resultados obtenidos en la misma ambos de elaboración sub grupal. A su vez los estudiantes deben informar en forma escrita semanalmente sus acciones en el centro de práctica y un análisis de las mismas.

En los casos de trabajos pautados se le presenta a los estudiantes con anterioridad a la realización de los mismo las dimensiones que

serán evaluadas y el valor de las mismas en la conformación del nota final.

Las dimensiones seleccionadas para los trabajos teórico práctico son:

- Pertinencia de la información que se expone (0,25%)
- Precisión conceptual en el desarrollo de las ideas que se presentan (0,30%)
- Bibliografía utilizada (0,20%)
- Adecuación a la organización textual propuesta (0,15%)
- Presentación formal (0,10%)

En investigación:

Proyecto:

1. Existencia de un objeto pertinente claro y definido (30%)
2. Lógica de exposición (30%) (coherencia y línea argumental; rigurosidad en el contenido teórico; articulación entre los contenidos y el objeto de estudio; definición teórica de los elementos centrales implicados en la pregunta de investigación)
3. Diseño de la investigación propuesto (30%) (pertinencia del diseño)
4. Aspectos formales (10%)

Informe de investigación:

1. Lógica de exposición (45%) (coherencia y línea argumental, rigurosidad en el contenido teórico, articulación entre los contenidos y el objeto de estudio, respuesta a la pregunta de investigación propuesta)
2. Exposición clara del proceso de investigación realizado y los resultados obtenidos (45%)
3. Aspectos formales (10%)

El equipo ha ido elaborando diferentes consignas y formas de procesar la evaluación (rubrica) que creemos pertinentes colocar a discusión.

Evaluación formativa en el primer año de Medicina

Martín Esteche, Paola Hernández

Resumen

Nuestra experiencia en Evaluación Formativa se desarrolla en el primer año de la carrera Doctor en Medicina de la Facultad de Medicina de la Udelar. Enmarcada en el Nuevo Plan de Estudios, implementado a partir de 2009, que procura formar «un egresado capaz de educar y educarse, crítico, preparado para trabajar en un sistema que priorice la atención primaria de la salud que es en definitiva lo que demanda nuestra sociedad» (Perfil del Egresado aprobado por la Asamblea del Claustro en 1995).

El primer año consta de dos ciclos: el Ciclo de Introducción a las Ciencias de la Salud (CIntro) y el Ciclo Básico Clínico Comunitario I, ambos cuentan con actividades curriculares, como Aprendizaje Basado en Problemas (ABP) y Trabajo de Campo (TC). Ambas son estrategias de enseñanza aprendizaje centradas en el estudiante, de trabajo grupal, donde el docente adopta un rol de «facilitador del conocimiento». Siendo nosotros docentes del Departamento de Educación Médica que actualmente nos desempeñamos en el CIntro.

En ABP trabajamos en base a un problema por semana, en la primer instancia se designan los roles de los estudiantes: moderador, asistente y secretario; se discute el problema, se plantean hipótesis y objetivos de estudio. Luego los estudiantes estudian esos objetivos y discuten por EVA. En la segunda instancia se realiza la validación e integración de los conocimientos y se realiza una tarea escrita sobre los temas tratados. Además deben realizar tres mapas conceptuales a lo largo del curso.

¿Cómo llevamos a cabo la evaluación formativa?

La evaluación continua es una evaluación llevada a cabo durante todo el curso y que tiene dos instancias formales de devolución al estudiante, una a mitad del curso y otra al final, no obstante hay un diálogo continuo.

La realizamos mediante una grilla de evaluación, que tiene 5 componentes: Análisis del problema, actitud frente a las propuestas de trabajo, actitud frente al grupo, roles, tareas y mapas. El análisis del problema es el eje de nuestro trabajo. Dentro de todos los ítems nombrados se evalúa la evolución a lo largo de las instancias.

Las metodologías de enseñanza-aprendizaje centradas en el estudiante y la evaluación formativa son pilares del Nuevo Plan de Estudios, y su implementación ha presentado múltiples desafíos y dificultades. Como desafío hemos trabajado en elaborar una herramienta que nos permita homogeneizar criterios y estandarizarlos, es así que surge la grilla de evaluación, que está en constante revisión por los docentes.

Contamos con múltiples instancias para poder trabajar estos elementos como son los Talleres de Evaluación organizados por el DEM, y las Supervisiones Formativas que son instancias semanales entre tutores donde se discute todo lo inherente a nuestra actividad.

Más allá que en estos 4 años creemos hemos ido consolidando esta experiencia, entendemos que hay múltiples aspectos a mejorar. Aún contamos con una grilla que tiene una gran subjetividad del docente, no existe continuidad en este tipo de evaluación a lo largo de la carrera y además contamos con la dificultad que no es ajena al resto de la Udelar donde la relación docente-estudiantil está lejos de ser óptima.

Experiencias de evaluación formativa en la Facultad de Derecho⁷

Javier López Portillo, Magdalena Casaña, Mariella Leles da Silva

Resumen

Las experiencias refieren a la evaluación de tipo formativo en dos cursos semestrales semipresenciales (año 2013), basadas en el desarrollo de trabajos colaborativos y el uso de las TIC.

En ambos se plantearon los objetivos pedagógicos al comienzo del curso, a través de: a) un documento conteniendo las condiciones de

⁷ *Materias: Métodos Alternativos de Resolución de Conflictos; y Mediación.*

cursado del curso y las estrategias pedagógicas a desarrollar, y b) un formulario de rúbrica con los ítems a valorar.

El nivel de aprovechamiento de la materia se mide en la mayor o menor dificultad de realizar la transposición práctica, y en el desarrollo de habilidades para interactuar en grupo y resolver problemas.

Al comienzo del curso se solicitó la conformación de equipos, de forma voluntaria, compuestos de 3 a 7 miembros. Se brindó capacitación previa en el uso de los recursos.

El material teórico se brindó a través de exposiciones (aulas presenciales), documentos (archivos, presentaciones) y videos especialmente preparados por el docente.

La evaluación formativa se llevó adelante utilizando: wiki, foros, tareas, cuestionarios, mapas conceptuales, y dramatizaciones .

Al finalizar el curso se pudo constatar:

- La importancia:
 - a) de la retroalimentación inmediata en la evaluación formativa;
 - b) del acompañamiento personalizado durante el proceso de aprendizaje, estimulando al estudiante a superarse.
- El escaso manejo del rol de la metacognición en el proceso de aprendizaje.
- Poca disposición y cultura de trabajo colaborativo, que se traduce en dificultades para interactuar.
- Autoevaluación: falta de entrenamiento; ausencia de criterios.
- Dificultad para asumir una modificación en el estilo de evaluación: los estudiantes están preocupados por la evaluación certificadora, en torno a la cual gira su actuación en el proceso de aprendizaje. Los apartamientos al modelo tradicional son vistos con desconfianza.

Conclusiones:

- Es preciso se difundan, estimulen, y generalicen nuevas prácticas de evaluación formativa.
- Es preciso capacitar no sólo a docentes, sino también a estudiantes, proporcionándoles las herramientas necesarias para comprender y consolidar el cambio.

Bibliografía

- Bain, K. (2007). Lo que hacen los mejores profesores universitarios; Universitat de Valencia.
- Bertoni, E. (s / f). «Evaluación en la Universidad», disponible en <http://www.cse.edu.uy/sites/www.cse.edu.uy/files/documentos/Bertoni_-_Evaluar_en_la_universidad.pdf>.
- Caamaño, C. (2009). «La evaluación de los aprendizajes en la universidad ¿Ejercicio de poder, regulador social o potenciador de los aprendizajes? Artículo publicado en la Revista Encuentros Uruguayos, Revista Interdisciplinaria de la FHCE, CEIU, pp. 260-287 (publicación arbitrada).
- Camilloni, A. (1998). «La calidad de los programas de evaluación y de los instrumentos que la integran», en Camilloni, A, Celman, S., Litwin E y Palou de Maté M. La evaluación de los aprendizajes en el debate didáctico contemporáneo, Buenos Aires, Paidós, pp. 67-92.
- Celman, S. (1998). «¿Es posible mejorar la evaluación y transformarla en herramienta de conocimiento?» en Camilloni, A, Celman, S., Litwin E y Palou de Maté M. La evaluación de los aprendizajes en el debate didáctico contemporáneo, Buenos Aires, Paidós; pp. 35-66.
- Litwin, E. (1998). «La evaluación: campo de controversias o un nuevo lugar para la enseñanza» en Camilloni, A, Celman, S., Litwin E y Palou de Maté M. La evaluación de los aprendizajes en el debate didáctico contemporáneo, Buenos Aires, Paidós; pp. 11-34.
- Litwin, E.; Palou, C.; Calvet, M.; Herrera, M.; Pastor, L. (2003. «Aprender de la evaluación» en Educación, Lenguaje y Sociedad, Vol. I n.º 1 (diciembre), Facultad Ciencias Humanas, Argentina; pp. 167-177.

Evaluación Formativa del Practicantado de Producción Animal de la Facultad de Veterinaria-Udelar

*Carlos Morón, Roberto Kremer, José Luis Repetto, Daniel Cavestany,
Claudia Borlido, José Passarini*

Resumen

El Orientado en Producción Animal corresponde al 5º año de la carrera de Dr. en Ciencias Veterinarias, se puede realizar en Paysandú (OPA-NORTE) y en el sur del país (OPA-SUR) integrado por Cursos Obligatorios (600 h), Cursos Optativos (100 h) y Practicantado (300 h). El Prac-

ticantado es el conjunto de actividades prácticas, vinculadas a la orientación, en el que el estudiante enfrenta los desafíos que le presentará el desarrollo de su futuro profesional. Particularmente para el Practicantado del OPA-SUR comprende las siguientes actividades: guardias de parto en establecimientos lecheros, pasantías en Ecilda Paullier (ATP), salidas con veterinarios, pasantías en Centros Experimentales, análisis de predios productivos, etc.

El seguimiento de las actividades se realiza mediante la confección de un Portafolio por parte del estudiante, el que es evaluado por el cuerpo docente.

El portafolio consta de informes de las actividades del practicantado incluyendo informes de salidas, informes de permanencia en predios, presentaciones de las transmisiones horizontales, monografías, etc. El portafolio se actualiza, recoge las sugerencias y correcciones de los docentes y da cuenta del progreso que puede realizar el estudiante a partir de diferentes experiencias, reflexiones a partir de conceptos teóricos y su relación con la práctica. Se valora tener el Portafolio completo, mejorar los informes definitivos de acuerdo a las correcciones realizadas, y agregar materiales pertinentes a la temática.

Por otro lado, el último viernes de cada mes se realiza una Transmisión Horizontal (presentación oral) donde los estudiantes presentan predios, casos clínicos, etc. También se incorporan temáticas de interés con invitados. En estas instancias participan los estudiantes, docentes y veterinarios del ejercicio liberal de la profesión, realizándose un intercambio de manera horizontal entre todas las partes.

A través del Departamento de Educación Veterinaria se hace un seguimiento de las actividades mediante la modalidad de entrevistas y encuestas a los estudiantes a mitad del semestre y al final del año. De estas evaluaciones, surge que los estudiantes consideran como una buena forma de evaluación los portafolios y los seminarios, consideran que estas estrategias permiten mejorar la comunicación oral y escrita de ellos como futuros veterinarios, así como mejorar algunos aspectos concretos: realizar citas bibliográficas y la futura realización de la tesis final de grado. Finalmente, el portafolio es entregado a cada estudiante para que le sea de utilidad como material de consulta sobre casos clínicos e informes sobre producción animal.

La incorporación de la evaluación estudiantil de la docencia en el IENBA

Selva Artigas

Resumen

Es tradicional que en las instituciones educativas la temática de la evaluación haga referencia principalmente a los procesos de aprendizaje realizados por los estudiantes en el campo de conocimiento. No obstante, en la actualidad, la evaluación educativa comprende un amplio espectro: la actividad de los estudiantes, de las instituciones, de los equipos docentes y no docentes así como de la actividad individual de los actores.

La experiencia a presentar en el marco de la modalidad «evaluación formativa» aborda como temática el proceso de evaluación de la docencia desde la perspectiva estudiantil, instalado en el Instituto Escuela Nacional de Bellas Artes (UR) desde el año 2008.

Atendiendo al carácter formativo, la evaluación de la docencia en el IENBA se planificó como un relevamiento evaluativo en dos momentos del desarrollo curricular: avanzado el primer semestre y finalizando el año lectivo. Esto posibilita una incidencia estudiantil en la marcha de los cursos al expresar su percepción sobre la misma y una oportunidad de reformulación o de afirmación de la trayectoria curricular por parte de los equipos docentes. La estrategia de evaluación integra aspectos cualitativos y cuantitativos de la percepción estudiantil a los efectos de otorgar sentido a los datos estadísticos colectados.

¿Qué aporta esta modalidad de evaluación a los docentes y a los estudiantes? Importa señalar que el intercambio docente-estudiantil, previsto en este proceso evaluativo al mediar el año, produce un cruce de fronteras entre el equipo docente y el grupo de estudiantes desde sus respectivos enfoques. Mientras que el primero actúa respondiendo a los lineamientos conceptuales predefinidos, sistematizados en los planes de estudio, los segundos actúan desde las experiencias vividas que impactan sobre sus expectativas, y que resultan coincidentes o no con la propuesta institucional. Estas percepciones desde ambos sectores deberán interpretarse mutuamente para conjugar la com-

presión, el ajuste y la adhesión a diversos aspectos de las trayectorias curriculares, resultante de esa interacción.

Bibliografía

- Camilloni, A. (1995) La evaluación de la docencia universitaria. Facultad de Odontología / UR. Papeles de Trabajo. Montevideo.
- Díaz Barriga, A. (2000) Evaluar lo académico. Organismos internacionales, nuevas reglas y desafíos. En: Evaluación académica. Teresa Pacheco Méndez y Ángel Díaz Barriga (coord.) CEU - UNAM. Ed. FCU. México.
- Errandonea, J. (1986) Aportes doctrinarios. Plan nuevo de Bellas Artes. En: Escuela Nacional de Bellas Artes – Proyección de su experiencia educacional. (1963) / UR. Montevideo.
- FEUU (1997) VIII Convención de la FEUU. Noviembre de 1997. Montevideo.
- Lukas, J. y Santiago, K. (2004) Evaluación Educativa. Alianza Editorial. Madrid.
- Umpiérrez, R. y sztern, S. (1999) Innovación Educativa en la Enseñanza Superior. En: Pedagogía Universitaria. Presente y Perspectivas. Cátedra UNESCO /UR. Montevideo.
- Universidad de la República (1997) La Universidad de la República en un tiempo de cambios. UR / 12 al 17 de mayo. Montevideo.

Experiencia de evaluación en el Instituto Superior de Educación Física

Mariana Sarni, Isabel Pastorino

Resumen

Establecer un proceso de evaluación formativa, en este caso centrada en los aprendizajes, supone inicialmente deslizar el centro de discusión de los profesores, los sistemas educativos y nuestra propia universidad, respecto a ¿cómo evaluar?, y en consecuencia ¿cómo mejorar el instrumento? para proceder a valorar los resultados de aprendizaje, recentrando el debate en el ¿para qué se evalúa? y a partir de allí ¿qué y entonces cómo?

Preguntarse sobre supuestos teóricos subyacentes y problemáticos, parece cuestión central para asumir y presentarse a la evaluación como formativa. Estos supuestos la asumen como construcción subjetiva y no aplicación objetiva, como compleja en formas y sentidos, y como estrategia fundamental a favor de la circulación del saber, del poder, y de la construcción de la autonomía de nuestros estudiantes y nuestros profesores.

La experiencia de evaluación del aprendizaje que compartiremos, es aquella que vehiculiza a la formación en la Práctica Docente del Instituto Superior de Educación Física.

Tiene como fundamento central a la evaluación como mediadora de los aprendizajes de la formación del profesor. En este marco entendemos práctica profesional de formación como una «subunidad organizativa y pedagógica con características idiosincráticas que la diferencian del resto de las asignaturas que componen el plan de estudio de la carrera» (Andreozzi, 1996, p. 23 y los aprendizajes como aquellos procesos en los cuales se maneja información que de forma flexible, pueda aplicarse a situaciones nuevas situadas, (Perkins, 1997).

Desde esta perspectiva el aprendizaje en un contexto incluye en su centro la naturaleza conflictiva de la práctica social. El carácter situado e histórico del aprendizaje, fuertemente influenciado por el contexto y la cultura, pone de relieve su carácter socialmente construido a través

de interacciones con otros y frecuentemente guiado por expertos en experiencias de aprendizaje compartido.

La evaluación en ese marco intenta superar su perspectiva eficientista, valorándose en consecuencia a modo de proceso que de manera intersubjetiva y mediando el debate sobre el conocimiento situado, permite a quien participa la (re) construcción del mundo a favor de su valoración por parte de los sujetos implicados.

Parte de la construcción/reconstrucción de los criterios que ofician de referentes sobre el referido profundo, *la buena práctica*, de la elaboración de documentos que son convocados para sostener/accompañar el proceso, y que triangulan/proyectan la intersección del saber propio de los practicantes, el saber producido en la escuela, el propiamente académico y el de los niños sujetos de aquellas prácticas, y finalizan con la valoración de parte de aquel grupo de los aprendizajes obtenidos en relación a aquellos que se planteaban como expectativas iniciales en relación con el referido original. Esto supone un proceso de calificación el cual también se instala como asunto negociado, sujeto a sentido.

Bibliografía

- Andreozzi *et al.* (1996). El impacto formativo de la práctica. En: Revista del Instituto de Investigaciones en Ciencias de la Educación. UBA. Año V, n.º 9. Buenos Aires, Miño y Dávila. 20-31.
- Perkins, D. (1997). La escuela inteligente. Del adiestramiento de la memoria la educación de la mente. Barcelona: Gedisa.

Ética y Profesionalismo en la enseñanza clínica de la ginecología: experiencia en simulación en estudiantes de grado y posgrado

Grenno, Curbelo, Cóppola, Briozzo

Resumen

El profesionalismo es una competencia transversal a toda la carrera de medicina y particularmente sensible en el área de la ginecología. La misma se debe enseñar y evaluar desde el inicio de la carrera y no al final de la misma. Por ello la Facultad de Medicina de la Udelar en el actual Plan de Estudios lo realiza desde el inicio.

Cambios en la currícula de la Facultad de Medicina de la Carrera de Doctor en Medicina inscriptos en el Nuevo Plan de Estudios (NPE), han determinado la inclusión de la enseñanza de la Ginecología en el primer semestre del tercer año, en el curso Ciclo Clínico Básico Comunitario.

En este curso la formación se realiza mediante diferentes metodologías de aprendizaje incluyendo discusiones grupales, clases expositivas y talleres de simulación (Paciente Simulado, Role Playing y simulación con maniqués ginecológicos). La grabación de los encuentros permitió a los estudiantes observar su desempeño y reflexionar acerca de sus fortalezas y debilidades. Estas actividades fueron seguidas de entrevistas a pacientes reales en el ámbito comunitario facilitando así la transferencia de lo aprendido y el desarrollo de la profesionalización de los estudiantes. La evaluación formativa en base a la observación directa del tutor y feedback constructivo, el portafolio como herramienta que promueve una actitud reflexiva de docentes y estudiantes y la evaluación final de tipo sumativo se lleva a cabo a través de un Encuentro Clínico Objetivo Estandarizado (ECO-E).

Se describe dos ejemplos de simulación, el primero es el Taller de Historia Clínica Ginecológica y Obstétrica, la metodología utilizada es la simulación mediante Paciente Simulado, videograbados, (actrices) y técnica de Role Playing. Esta metodología permite al estudiante adquirir actitudes, habilidades específicas y definidas en el interro-

gatorio, logrando así habilidades comunicacionales que lo preparen para los encuentros con pacientes reales. Se pretende que el estudiante adquiera competencias para actuar profesionalmente frente a los pacientes, salvaguardando la ética en la enseñanza y mejorando la calidad de la misma. Se aborda en forma transversal la ética médica, autonomía, derechos de los pacientes, secreto profesional, confidencialidad, consentimiento informado.

La segunda experiencia consiste en el taller de Examen Físico Ginecológico y Obstétrico se realiza mediante la simulación con maniqués pélvicos.

En ambas instancias se realiza una evaluación formativa con devoluciones por parte de los docentes y pares que promueva la reflexión y el aprendizaje a la vez se prepara al estudiante para el trabajo en equipo.

Se pretende con los talleres de simulación que el estudiante se involucre con los procesos de aprendizaje, sepa qué es lo que se quiere enseñar, cómo se quiere enseñar, cuáles son los objetivos de aprendizaje que se buscan y cómo va a hacer para lograrlo.

Se utilizan listas de cotejo facilitando la autoevaluación, evaluación entre pares y docente.

Esta primera experiencia de simulación en ginecología en el NPE, culminará con una encuesta de opinión a los estudiantes, lo que nos servirá para poder planificar próximos cursos.

Experiencia en la Facultad de Ciencias Económicas y Administración⁸

Alba López Murissich⁹

Resumen

Esta experiencia tuvo lugar en el marco de un Proyecto¹⁰ a cargo de Selva Artigas.¹¹ Al comenzar el curso informé a los estudiantes de las actividades que íbamos a realizar. Una de las técnicas de aprendizaje que más utilizamos independientemente de la Evaluación Formativa (EF), es el método de estudio de casos, tratando que el estudiante analice y tome decisiones sobre situaciones semejantes de la vida real que se describen en un «caso de estudio». Esta forma de trabajo implica la formación de subgrupos y se desarrolla en tres etapas: Una primera de preparación individual, luego una de discusión grupal, para culminar con un comentario en sesión plenaria. De esa manera se logra involucrar al estudiante en forma individual, para luego intentar lograr un aprendizaje comunitario.

Algunas técnicas manejadas a lo largo del curso a efectos de desarrollar un proceso de EF fueron:

En la primera clase nos presentamos, indago el porqué de la elección de la carrera, cuántos trabajan, en qué, otros estudios, con quién viven, de dónde son. De esta forma se produce un acercamiento que luego se profundiza a lo largo del Segundo Encuentro de la Red UAE 2014 ¿De qué Evaluación Formativa Hablamos?

semestre. Luego realizamos un caso de estudio sencillo, que les permite interactuar y comenzar a trabajar en subgrupo.

.....
8 *Curso “Organización y Sistemas I” – Escuela de Administración - 2º año de la carrera de Técnico en Administración - Primer semestre de 2012*

9 *Prof. Asistente Interino - Facultad de Ciencias Económicas y de Administración - Unidad Académica Administración y Gestión de las Organizaciones – Unidad Curricular Procesos y Sistemas de Información – albalopezmurissich@gmail.com.*

10 *“La Evaluación Formativa como potenciadora de los aprendizajes on line” Octubre-2011 a Octubre-12.*

11 *Prof. Agr. de la Unidad de Apoyo a la Enseñanza de Facultad Ciencias Económicas y de Administración*

- Evaluación diagnóstica sobre contenidos básicos del curso.
- Controles de lectura.
- Entrega de preguntas para resolución.
- Lluvia de ideas sobre conceptos claves antes de comenzar los temas.
- Entrega de síntesis de lo hecho en subgrupo en cada clase.
- Utilización del correo electrónico para dudas y corrección de casos prácticos extras.
- Entrega de prueba de revisión no aprobada para que la corrijan, luego se analiza.
- Solicitud de preparación de temas teóricos, para presentación oral en clase.
- Entrega al final de clase de una pregunta a contestar de 1 o 3 minutos para ver que entendieron o que no, para retomarlo a la clase siguiente.

Las actividades se fijan con anticipación e implican necesariamente el disponer de tiempo extra para su realización, tanto de parte del estudiante como del docente, quien tiene que dar devoluciones de lo solicitado.

La mayor dificultad que se nos presentó se relacionó con el tiempo que disponemos para llevar adelante el curso, que es breve y su cronograma de clases inflexible.

Lo bueno de poder aplicar la EF es que produce un feedback que contribuye a regular los procesos de enseñanza-aprendizaje, desarrolla habilidades meta cognitivas en los alumnos, permitiéndoles tomar conciencia de sus propios recursos, lo que ayuda a promover su propia autorregulación.

Para finalizar, considero que en líneas generales pudimos desarrollar las actividades, agregando algunas clases extras. Los estudiantes se percibieron motivados, respondiendo a la mayoría de las propuestas, logrando como resultado final ser parte activa en su proceso de formación. Creo que es necesario incorporar en forma permanente la EF, pero dándole los tiempos que requiere, ya que es útil al estudiante, y además se estaría cumpliendo con la Ordenanza de Grado que la establece (Art. 37).

Bibliografía

- Camilloni, A. y otros. (1998). *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Ed. Paidós, Buenos Aires.
- Ordenanza de estudios de grado y otros programas de formación terciaria - 2011 – Udelar - Resol. N° 3 CDC 2-8-2011 - Resol. N° 4 CDC 30-8-2011 - D.O. 19-9-2011.

La Evaluación entre pares. Experiencia del ciclo introductorio. Facultad de Medicina

Arrillaga, Abelenda

Resumen

La implantación de un sistema de calidad total en el ambiente universitario pasa obligatoriamente porque todos los individuos del colectivo universitario tengan una actitud positiva hacia la mejora continua, lográndose así la idea de calidad como cultura (Manso, 1997).

Desde el año 2010 se ha implementado una experiencia de co-evaluación entre los docentes de Acción Tutorial del ciclo introductorio (primer año) . Dicha decisión partía de la realidad de contar con un número importante de docentes que no tenían experiencia en la enseñanza universitaria por lo que se veía imperioso el realizar una Evaluación Formativa que aportara datos del acontecer y que fuera una aporte para la tarea docente.

¿Qué dificultades existen a la hora de realizar la evaluación docente?

- La dificultad asignada a la tarea de evaluar
- El no retorno de la información a los participantes
- La ausencia de explicaciones.
- La falta de actividades de autoevaluación

Existen problemas como:

- La ausencia de una clara comprensión acerca de lo que es un buen docente
- Valoraciones basadas más que nada en productos que poco tienen que ver con la cotidianeidad de los docentes.

- Existe poca diversificación de instrumentos para llevar a cabo la evaluación de la docencia. (Murillo 2008)

¿Qué es la evaluación entre pares?

- La coevaluación se refiere a la evaluación entre pares, entre iguales. Es aquella forma de evaluación donde los evaluadores y evaluados intercambian sus roles alternativamente.

¿Qué ventajas se podrían obtener de este tipo de evaluación?

- Media el dialogo entre docentes.
- Se comparten situaciones similares, pudiendo brindarse apoyo en asuntos vinculados al quehacer docente.
- Enriquecimiento del evaluador a través de la observación de otra práctica docente y su posterior feedback.
- Los evaluadores pares identifican potencialidades y necesidades en sus colegas.

¿Cuál es el momento de su aplicación?

- La bibliografía considera que la co evaluación se debe realizar dos veces en el semestre académico

La Experiencia en Facultad de Medicina

- Se realiza desde el año 2010.
- Utiliza una Grilla de observación, que ha variado a lo largo del tiempo.
- Consta de 2 instancias presenciales en un mismo problema en la que se observa el desarrollo normal de una instancia de abp, donde el observador no participa activamente de la clase.
- Se promueve un Feedback entre el observado y observador y la realización de sugerencias.
- Se realiza un informe en base al desempeño del rol y la instancia de retroalimentación que deberá enviar al tutor observado.

Bibliografía

Aparcana F; Co-evaluación del desempeño docente. [Sede Web]. Lima, Perú. 2011. Fecha de acceso: 12 de setiembre de 2013. Disponible en: <<http://www.usmp.edu.pe/>

publicaciones/boletin/fia/info82/articulos/evaluacion1.pdf?TB_iframe=true&height=600&width=550>.

Correal M; Beltrán M. La evaluación docente en educación superior: uso de instrumentos de autoevaluación, planeación y evaluación por pares. [Sede Web]. Universidad Nacional Autónoma de México. Instituto de Investigaciones sobre la Universidad y la Educación. Fecha de acceso: 12 de setiembre de 2013. Disponible en: <http://riied.uniandes.edu.co/archivos/Taller_3/Mario_Rueda_Marisol_Diego.pdf>.

«Temas de enseñanza», colección publicada por el Prorrectorado de Enseñanza, reúne conferencias, estudios, ponencias y debates generados en jornadas institucionales. Se propone difundir información, reflexiones y conocimiento acumulado en distintas áreas del campo educativo.



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Red de Unidades de
Apoyo a la Enseñanza

ISBN: 978-9974-0-1214-1