



## Evaluación como herramienta de impacto: **¿Es esto beneficioso o dañino?**

**Nick Saville, University of Cambridge Esol,**  
Procesos de Investigación y validación, Inglaterra

**Segundo Día  
CONFERENCIA**



La enseñanza y el aprendizaje se ven influenciados por las prácticas de evaluación, pero a medida que se van diseñando los exámenes para cumplir con políticas y también para dar los indicadores de rendición de cuentas, se hace aún más importante que estos exámenes den evidencias de que las consecuencias sociales de estos exámenes traen beneficios, y hay que tratar de que estos beneficios se maximicen. De igual manera es importante que nos preguntemos de qué manera se pueden evitar las consecuencias dañinas o se pueda mitigar su impacto; por lo tanto esta charla tiene que ver con las consecuencias de la evaluación de la enseñanza del lenguaje y de su impacto.

Se necesita una mejor forma para entender los complejos factores que determinan si el impacto de los exámenes de idiomas va a tener consecuencias positivas o negativas de acuerdo con las polí-

ticas fijadas. Estas consideraciones son aun más relevantes cuando se introducen proyectos de reforma a nivel nacional. Para lograr esto, el impacto tiene que ser una parte integral al desarrollar los sistemas de prueba y enseñanza del idioma, y estoy hablando del diseño, un concepto que es parte de la evaluación dentro de este contexto social. Pero primero veamos las definiciones de washback y de im-

pacto: washback ha sido definido como el efecto de hacer las pruebas en la enseñanza y en el aprendizaje, y el impacto es un aspecto más amplio de ese efecto.

En 1983 en un trabajo sobre washback, se sugirió que este término nos ayuda a explorar el papel de las pruebas en lo que tiene que ver con el educando individual, las actitudes y el comportamiento del maestro, la selección del material didáctico y demás; y desarrollaron entonces 15 hipótesis de washback basadas con quien pudo verse afectado



Nick Saville

DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA



## ¡Evaluar es valorar!

Foro educativo nacional sobre la evaluación de los aprendizajes en el aula



DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA

## Evaluación como herramienta de impacto: ¿es esto beneficioso o dañino?

por estas pruebas, incluyendo la enseñanza, la velocidad, el contenido, la secuencia, la profundidad y el número de personas involucradas. Este campo del washback con retroalimentación nos habla del impacto de las pruebas en un continuo, que va desde un impacto negativo en un extremo a un impacto neutro en el centro y un impacto positivo en el otro extremo. El impacto negativo ocurre cuando la evaluación se basa en una definición muy restringida y por lo tanto faltan muchas cosas por enseñar, limitando el aprendizaje y dando la enseñanza en una forma muy estrecha. El washback positivo ocurre cuando los procedimientos de evaluación promueven buenas prácticas, y llevan a experiencias de aprendizaje positivas, por ejemplo dar una mejor motivación a los educandos o una mayor transparencia de lo que significa el aprendizaje. Sin embargo, ha quedado claro en el transcurso de los últimos 15 años, que hay muchas variables que influyen en el washback que podemos encontrar dentro del contexto educativo, y que la explicación de causa y efecto no es la adecuada y que por lo tanto necesitamos ver esto en una forma mucho más profunda y compleja. Varios investigadores han desarrollado unos modelos de washback más complejos y han sugerido la forma como se puede utilizar en la agenda de investigación de una forma más sistemática. Las publicaciones del Bailey son muy conocidas y muy informativas al respecto. Por ejemplo Bailey en 1996 propone un modelo de tres partes que se concentran en las tres P: Participantes, procesos y productos. Contextualiza el modelo en términos de 5 dimensiones: especificidad, intensidad, duración, intencionalidad y valor y el sugiere que los aspectos de la enseñanza y el aprendizaje pueden verse influenciados por un examen y que los factores que facilitan la enseñanza, en el washback están relacionados con la prueba, con el estado y con los actores que participan.

Creo que el modelo más completo de washback es el de Green. Su modelo cubre la direccionalidad, la variabilidad y la intensidad. Entonces para resumir, el efecto del washback es más intenso o tendrá los efectos más poderosos en los procesos de enseñanza y aprendizaje, cuando los participantes vean la prueba como un reto y los resultados como algo importante; en el caso de las pruebas estatales, hay una fuerte creencia de que las pruebas de examen van más allá de este contexto y un reconocimiento de que las pruebas de idiomas pueden tener un impacto en las oportunidades de vida de los individuos, por ejemplo en el acceso a las oportunidades de educación o de empleo. También puede tener un impacto en los sistemas educativos, los cuales son parte de la sociedad en su conjunto, por ejemplo cuando se utilizan estas pruebas para tomar decisiones importantes sobre la planeación curricular o la asignación presupuestaria para las escuelas, o la política de inmigración o licencias para profesionales de la salud como médicos y enfermeras, los intereses comerciales como por ejemplo los editores que preparan el material para pruebas y las instituciones que dictan cursos para preparar para estos exámenes u otras actividades o negocios relacionados con esta evaluación. Entonces, los efectos o consecuencias de los resultados de una prueba a nivel macro, a nivel de la sociedad, es lo que denominamos impacto, el washback o efecto colateral es apenas un aspecto del impacto relacionado con los micro contextos como el aula de clase o el centro educativo. El impacto a nivel macro, con base en el modelo desarrollado por Messick, tiene que ver con los as-

“El concepto básico de las escuelas incluyentes es que todos los estudiantes, todo tipo de estudiantes trabajan en los mismos grupos con un solo docente.”





DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA

pectos de las consecuencias de la validez y conjuntamente con los temas de imparcialidad y ética. Esa validez ha sido objeto de mucho interés en los últimos años y la mayoría de las pruebas de idioma reconocen que es un impacto complejo que requiere de una investigación sistemática.

Para mí un problema que tiene que ver con la investigación es precisamente la relación entre el contexto micro y macro como espero poderles demostrar más adelante. Ahora, el entender estos fenómenos incluye una serie de interrelaciones entre los actores del proceso educativo o de los diferentes participantes como los llamó Bailey. En nuestro caso del Cambridge ESOL, estos actores o participantes son quienes participan en el mundo de la enseñanza del idioma inglés y en los exámenes del idioma del inglés como idioma extranjero en distintos países del mundo. La prueba, la forma como está construida la prueba y, lo que es más importante, los resultados que genera, a su vez son utilizados en la sociedad por los diferentes actores. Ahora, los educandos y los docentes son los que se ven más impactados por este examen, pero también otros actores se ven impactados en diferentes formas, incluyendo los padres, los empleadores, los editores, los examinadores, etc.

Es evidente que las diferentes pruebas tienen diferentes actores o participantes. Para analizar el impacto es importante tener la retroalimentación de estas diferentes voces, el consultar con la comunidad de los diferentes actores o participantes es una forma de determinar su alcance e influencia. Vamos a basarnos entonces en la experiencia que Cambridge hizo.

Antes de hablar del enfoque de Cambridge quisiera hablar de la importancia de entender la forma en que los niveles micro y macro pueden interactuar a nivel de los procesos educativos, y para hacerlo vamos a ilustrarlo con dos casos

que han sido de interés tanto de los Estados Unidos como en el Reino Unido en años recientes, y que demuestran el problema de los incentivos perversos y resultados negativos en las evaluaciones; ambos casos tienen que ver con políticas educativas específicas a nivel nacional e ilustran la ley de consecuencias no intencionales a raíz de decisiones tomadas. ¿Que es esto de la ley de consecuencias no intencionales? En su trabajo en 1936 el sociólogo Robert aplicó un análisis sistemático a la problemática de las consecuencias no anticipadas de la acción social. Para Robert estas consecuencias no intencionales son los resultados de acciones en una acción que no se limita a lo que el actor originalmente tuvo la intención de hacer. Entonces así surge esta ley de consecuencias no intencionales que nos dice que cualquier acción produce una consecuencia no intencional, o sea efectos secundarios. Hay tres tipos de estas consecuencias no intencionales a las cuales hacemos referencia: a) una consecuencia o un beneficio positivo no esperado; b) un efecto perverso o negativo (que es lo que nos interesa en este caso) y c) una fuente de problemas como resultado de la ley de Murphy que dice que si algo puede fallar, falla. Entonces, hay muchos casos que podrían ser citados para ilustrar esta teoría a nivel de la medicina, o del derecho que por razones de tiempo no voy a entrar en detalle. Sin embargo al analizar estos fenómenos es obvio que las causas de estas consecuencias no intencionales, muchas veces están relacionadas con la complejidad inherente a los sistemas dinámicos, que por una parte son resultado del sistema social que responde a cambios, y por otra parte a cambios en el entorno más amplio circundante, y estos cambios no se pueden predecir fácilmente y pueden verse afectados por condiciones aleatorias; esto también es cierto en los sistemas educativos que a su vez son complejos y dinámicos.



## ¡Evaluar es valorar!

Foro educativo nacional sobre la evaluación de los aprendizajes en el aula



DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA

## Evaluación como herramienta de impacto: ¿es esto beneficioso o dañino?

Entonces a veces observamos que la interacción entre una política educativa, los incentivos que se crean para ciertas acciones y las características individuales de personas involucradas, llevan a consecuencias no intencionales y no previstas, incluyendo impactos negativos en los resultados del aprendizaje, y estas consecuencias negativas pueden considerarse que son un resultado de los incentivos perversos de la política. Quisiera ilustrar esta teoría con algunos ejemplos, una variable de la ley de las consecuencias es lo que se conoce como la ley de Campbell en los Estados Unidos o la ley de Goodhart en el Reino Unido, llevan el nombre de los autores que lanzaron estas teorías; Campbell que fue el primero decía que una vez que un indicador social o económico es el objetivo para efectos de una política económica o social pierde la información que le permite desempeñar este papel, en otras palabras que el indicador se convierte en el blanco por su propio derecho y esto tiene el objetivo de cambiar lo que se va a medir. Ahora, en ambos continentes esta ley señala el peligro de que se produzcan consecuencias negativas, cuando las pruebas diseñadas para mejorar rendimiento también se utilizan para monitorear los niveles educativos; entonces en el Reino Unido el sistema educativo allí ha sido pensado por 30 años de reformas educativas que podrían llamarse reformas dirigidas por la evaluación misma, en donde se ha explotado el efecto de esta evaluación en las escuelas. Sin embargo los investigadores en el Reino Unido han mostrado que las consecuencias de las políticas han sido contrarias a lo que se tenía la intención de obtener con esas políticas; en la Universidad de Londres, por ejemplo, han observado que el uso de los resultados de los exámenes públicos para monitorear el rendimiento de los centros educativos, cumple con esta ley, y que ha tenido el impacto de limitar el currículo y enseñar solamente para la prueba, sin lograr mejorar el rendimiento de los educandos que era la intención inicial. También

hemos detectado que una consecuencia del llamado decreto de que ningún niño debe quedarse rezagado, es que pierde su valor como indicador de los estándares y entonces pierde de vista su objetivo original. Entonces, este efecto de washback que vemos en este caso, o sea, el limitar el currículo, no equivale a expugnar cual es el resultado de la evaluación, sino que tiene que ver con el impacto más amplio y con el uso e interpretación de los resultados de la prueba en la sociedad por parte de los diferentes actores.

Los efectos perversos nos ilustran de que manera las políticas educativas pueden crear una tensión entre los objetivos educativos a nivel micro, la enseñanza y aprendizaje en el centro educativo y el nivel macro. Estos ejemplos también nos ilustran la importancia de que los educadores y







DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA

quienes hacen la evaluación, trabajen de la mano juntos para hacer frente a este dilema, para prever los potenciales impactos, ya que los evaluadores son parte de la agenda más amplia y pueden monitorear lo que ocurre en la práctica, tanto macro como micro, y si se utilizan métodos de investigación apropiados tanto cualitativos como cuantitativos y también para tomar acciones remediales, según sea necesario, de forma oportuna. De esta manera surgen los siguientes interrogantes, desde mi punto de vista: ¿Qué papel pueden desempeñar los exámenes? ¿Qué papel puede desempeñar el examinador? ¿Cómo se puede comprobar que estas pruebas sean las correctas de acuerdo con el propósito? Haciendo la investigación necesaria para ver qué está pasando dentro del contexto y tomando las acciones necesarias para mitigar los efectos negativos, en caso de que se presenten consecuencias negativas. Y aquí es donde voy a ilustrar nuestro enfoque a esta investigación y este impacto en Cambridge ESOL. En Cambridge nosotros hemos adoptado y desarrollado un enfoque sistemático para hacer una investigación del impacto como parte integral del proceso de desarrollo y validación de pruebas. Milanovic y Saville en un trabajo de 1996 proponen cuatro elementos del impacto de las pruebas, en un intento por investigar el impacto en forma integral de las prácticas de trabajo de Cambridge ESOL como consejo examinador a nivel internacional. Estas cuatro variables se expresan de la manera siguiente:

1. Máxima uno: Utilizar un enfoque racional para el desarrollo de la prueba.
2. Máxima dos: Apoyar a los actores en el proceso de la prueba
3. Máxima tres: Dar una información transparente

4. Máxima cuatro: Monitorear y evaluar, o sea recoger todos los datos relevantes para hacer el análisis de seguimiento.

Y estos fueron diseñados para ser cortos, para incluir los principios básicos y para guiar la toma de decisiones, para funcionar como recorders de estas máximas y de lo que es más relevante y más importante. Y quisiera revisar esto muy brevemente. La Máxima uno determina que se tiene que hacer una adecuada planeación utilizando un modelo racional para el desarrollo de pruebas. En los exámenes de Cambridge esto se define como un modelo de proceso en donde estas máximas son la parte central. Luego la validación de las pruebas es vista como una parte integral de este modelo del proceso e incluye el recoger de forma sistemática los datos para validar los exámenes una vez que se han hecho y esto tiene que ver con el concepto de que la validación es la acumulación, es la evidencia de la valoración. Este ciclo de desarrollo tiene que ver con las consideraciones del propósito de las pruebas, incluyendo el anticipar o prever de que manera podrían utilizarse estas pruebas, qué tan relevantes y útiles pueden ser en términos de sus consecuencias sociales y en términos de las implicaciones de valor, y sus posibles efectos, incluyendo los efectos secundarios. Es crítico que sea cíclico y reiterativo el conocimiento y la experiencia entre en las diferentes etapas, y se dé una retroalimentación, se continúen evaluando los exámenes y la administración de los mismos.

La máxima dos requiere que los actores sean apoyados en este proceso por los examinadores, esto es crucial cuando diseñamos nuevos exámenes, pero también un factor muy importante en los procesos de validación y para alcanzar el





impacto esperado. Siempre ha habido una relación directa entre los actores de Cambridge a nivel mundial y eso aparece en las publicaciones donde se muestra cómo han evolucionado los exámenes en los últimos años. Los cambios en metodología dan oportunidades para consulta directa con los participantes, como los profesores, los escritores, etc. El contacto directo con los centros a nivel mundial y las organizaciones que eligen nuestras calificaciones de exámenes, hacen que nosotros rutinariamente recojamos datos y obtengamos la retroalimentación sobre estos exámenes, pero a esto debe agregarse los análisis de los participantes, la opinión de los participantes y podemos incluir y analizar los efectos negativos. Las técnicas de investigación cualitativa son igualmente importantes para esto. Nuestros programas de seminarios, que aparecen en nuestros portales, han hecho nuestros canales de comunicación mucho más flexibles y más interactivos que en el pasado. Entonces, una misma persona está participando de formas diferentes, tanto como usuarios de los exámenes como también ayudando a diseñar los exámenes y asegurando que estas pruebas o exámenes sean valiosos para ellos y útiles en sus propias vidas o experiencias.

En la máxima número tres, el proporcionar información minuciosa, transparente y útil, es muy importante para la máxima número dos, y se hace utilizando un tipo apropiado de documentación, dando panfletos, libros, y seminarios. Esta máxima se refiere a dar aclaraciones sobre problemáticas relacionadas con la lengua y sus exámenes. La transparencia es como debemos comunicar los resultados de estos exámenes y la necesidad de la evaluación para los usuarios y participantes.

La máxima cuatro se refiere a la recolección de datos importantes y los requerimientos para analizarlos y evaluar los hallazgos. Una vez que estos exámenes son totalmen-

te operacionales, cuando los exámenes se utilizan en este macro-contexto, debemos agregar datos del desempeño de los candidatos. Ese análisis debe ser utilizado para evaluar el uso de los exámenes, para ver si hay revisiones que deben hacerse, y si se debe volver a la fase de planificación y comenzar el ciclo de nuevo. Adoptando este modelo el proveedor de exámenes puede estar en contacto con las consecuencias que son el resultado de contar con los elementos apropiados para la recolección de estos datos, metodologías apropiadas de investigación, estudio de casos, recolección de datos, etc.

Entonces, como conclusión, la investigación del impacto la experiencia durante los 15 años sugiere que los que suministran exámenes y los que planifican la educación deben trabajar juntos, que para asegurarse que los sistemas de evaluación alcancen sus objetivos deben analizar el impacto e integrarlo a un modelo operacional para su validación deben establecer una metodología para analizar el impacto y que los hallazgos de estos estudios de impacto deben distribuirse. La investigación del impacto realizada por una organización refleja la importancia de la planificación de la educación a través de estrategias de validación, y esa investigación de impacto sirve para que crear futuras acciones políticas de evaluación se pueden desarrollar para satisfacer las necesidades de la educación actual. Muchas gracias, por su atención.

*Desde los 1970, todos los maestros y docentes hacen una maestría y esto también es cierto para los docentes de primaria, básica y preescolar.*



La estandarización de pruebas en la lengua no afecta el impacto de los intereses de los participantes, estudiantes en su contexto socio-cultural





DÍA

DÍA

Conferencias

Videoforos

Experiencias

Panel 2

DÍA



Los exámenes internacionales tienen un efecto en el contexto local, si yo creo que cada vez que nosotros intervenimos en un sistema, nosotros cambiamos este sistema de cierta forma, entonces, utilizando los exámenes internacionales producimos algo diferente algo que no estaba ahí. La pregunta es si esto puede ser positivo y tener efectos beneficiosos o puede ser negativo; yo creo que como ha sido ilustrado, lo que le interesa a los organismos internacionales, no es llegar al medio ambiente de alguien y tratar de imponer soluciones o tener dudas preguntas; obviamente esto no es correcto ya que no tiene en cuenta las opiniones y actitudes locales, las creencias locales, que son tan importantes en cualquier tipo proceso de educación; entonces trabajando en colaboración, los beneficios de la participación internacional debe darse en un proceso de doble vía donde el socio internacional, reconoce la necesidad del socio local pero el socio local entiende cuáles son los beneficios adicionales que se pueden obtener al trabajar internacionalmente por ejemplo, en el caso de la educación por idiomas, y si nosotros estamos aprendiendo el Inglés como lengua extranjera realmente no es una materia académica, es algo necesario, y el objetivo final es lograr utilizar este idioma en un contexto práctico, para trabajo, para estudio, para viaje, para comunicarse con las personas en otro país como una manera de comunicación; como estudiantes debemos tener objetivos útiles que son un punto de referencia en nuestras expectativas a nivel internacional, no podemos salir de nuestro país creyendo que hablamos Inglés para descubrir que no sabemos casi nada aunque han tenido buenas

notas por sus profesores. Para la enseñanza de idiomas necesitamos ciertos puntos de referencia contra los cuales nosotros podemos medir nuestro progreso si es real o no y es ahí donde los organismos internacionales pueden ayudarnos a definir lo que es aceptable y cuáles son los niveles. Es importante tener un marco de trabajo que vaya más allá de los países, que permitan a las personas que ofrecen la educación, a profesores, y a examinadores, tener un diálogo y hablar más o menos de las mismas cosas, entender los niveles, y entender los desempeños.



Quando habla de efectos secundarios en evaluación a que se refiere. Un factor importante precisamente en la evaluación son los padres de familia. De qué manera ustedes trabajan con ellos en este aspecto?



Los efectos secundarios, cuáles son? En el caso de la educación en idiomas la mayoría de las personas consideran lo importante cuando se aprende una lengua extranjera es poder hablarla, ahora el examen al final del año se hace solamente en lectura y es frecuente que los profesores enseñen únicamente lectura y por consiguiente el elemento importante que se quisiera encontrar sería el hablar, pero no se hace esto no porque ellos consideren que no es importante sino porque consideran que deben ayudar a los estudiantes a pasar los exámenes y concentran su atención más en la enseñanza de la lectura; esto no es una consecuencia deseada pero es la que obtenemos como efecto secundario y este es un típico efecto, una característica del washback. Estos son el tipo de efectos secundarios que van más allá del salón de clase, trataré de ilustrar esto con la idea de las mesas directivas en el Reino Unido.





DÍA



DÍA



Conferencias

Videoforos

Experiencias

Panel 2

DÍA



## Evaluación como herramienta de impacto: ¿es esto beneficioso o dañino?

En los países donde no tienen esto tal vez sea difícil entender lo que quiero decir pero en el Reino Unido el gobierno quiere medir la calidad de los colegios por los resultados que ellos obtienen en sus exámenes, en los colegios asumen pues que si un colegio es capaz de que sus estudiantes tengan notas elevadas o altas, en sus exámenes este debe ser un buen colegio y por consiguiente de un colegio que tiene notas bien altas se puede poner la primera parte de la clasificación, como primero en la clasificación y así sucesivamente.

Y es importante para los padres en cuanto a qué colegio mandan a sus hijos, pero esto no es categórico porque ahí se falla en tener en cuenta el efecto de la enseñanza en los resultados; si tenemos estudiantes que vienen de una área socio-económica baja es muy difícil llevarlos a los altos niveles de los colegios por los resultados, pero los profesores pueden estar haciendo un buen trabajo en enseñarles a ellos, esto es aquí donde muchas personas, y yo veía que la política que utiliza esta lead-table para medir la eficiencia de los colegios están perdiendo, no está teniendo ciertos datos de información sobre la calidad de la educación y tienden a fomentar en los profesores a insistir más en los exámenes y no enfocarse en un currículo más amplio como el del aprendizaje de los estudiantes en una manera positiva. En este caso no se le exime al mismo que se tiene la falta, es la política la que hace uso de los resultados y si los participantes actores en los gobiernos no toman de decisiones que hace que esto suceda, es que no han querido hacerlo, por consiguiente es una consecuencia, que siempre surge a raíz de esto.



Cuánta importancia le da Esol a la fluidez de la lectura oral entre los criterios de logro, tiene un estándar al respecto.



Entiendo que esta pregunta es sobre habilidades de lectura. Si, obviamente es muy importante para que los exámenes expliquen lo que una persona que aprende puede hacer cuando ha aprobado un examen, porque el examen mismo es un mito de muestreo de lo que el estudiante ha aprendido. Lo que los padres y lo que las otras personas deben hacer es ver si para esta persona que aprende o este niño, puede leer cosas que están relacionadas con las actividades de la vida real; debemos saber qué podemos hacer con el aprendizaje, pueden leer pequeños textos con 100 palabras, o artículos largos o en los periódicos o trabajos académicos; para establecer esto claramente necesitamos cierta teoría sobre lectura y sobre la forma de hablar para que haya una progresión desde el nivel básico al máximo y debemos desarrollar uno que refleje esta teoría para que adecuadamente podamos interpretar el resultado del examen en términos de lo que la persona pudiera hacer en el mundo real, fuera de la escuela o fuera del aula, y obviamente esto es lo que los buenos proveedores de exámenes deben hacer al desarrollar los exámenes, y como lo indiqué en mi charla explicar el significado de los resultados de los exámenes a los usuarios en términos que ellos puedan entender, no en términos que los evaluadores de exámenes utilizan y estos son los métodos o el marco de trabajo que utilizamos en el programa de inglés. Esto es muy útil porque nos da las bases para interpretar el significado de los niveles y dar a los profesores y otros actores claras indicaciones de lo que deben hacer para pasar de este nivel al



**¡Evaluar es valorar!**

Foro educativo nacional sobre la  
evaluación de los aprendizajes  
en el aula



Nick Saville

próximo y al próximo de una manera sistemática.

Muchas gracias al profesor Nick Saville de Cambridge Esol quien nos habló de importancia de la retroalimentación en el proceso de evaluación de los aprendizajes.

DÍA



DÍA



Conferencias


Videoforos

Experiencias

Panel 2

DÍA



 **Vínculo en internet**



Consulte el microsítio en internet del  
Foro de Evaluación de aprendizajes