

EVALUACIÓN: DEL AULA AL EXAMEN

Anna Golubiátnikova Merechénkova
EOI Girona
agolubia@xtec.cat

... no es lo mismo aprender a conducir que aprender a sacarte el carné. Esto obliga a repensar las técnicas de evaluación para que sean más exhibiciones de conocimiento que no pruebas para provocar el error. Que induzcan más a tener competencias que a saber aparentarlas.

Carles Capdevila
Si miras el retrovisor, que se note
"La vida que aprendo", 2017

LA FELICIDAD DE LOS DESGRACIADOS

Represento aquí seguramente uno de los grupos de redactores más desgraciados. Sabemos perfectamente que la validez y la fiabilidad de las pruebas que elaboramos tienen que ser avaladas por **unas estadísticas serias y sólidas**. Pero trabajamos con una de esas lenguas que en argot profesional se llaman "pequeñas". **No tenemos la suerte de contar con la posibilidad de realizar un pilotaje masivo** para poder analizar sus resultados basándonos en datos creíbles.

En Cataluña no pilotamos nuestras pruebas fuera de nuestra comunidad. Tenemos un número muy limitado de alumnos cuyos resultados podemos analizar.

Pero con todo mi respeto por las estadísticas me gustaría confesar que los alumnos y la lengua que enseñamos tienen para mí mucha más importancia.

Así que somos muy afortunados, mejor dicho afortunadas. Considero que con todo mi respeto por las estadísticas (que no tenemos datos muy fiables) **tenemos muchas otras ventajas** que nos permiten llegar en cierto sentido más allá de las estadísticas.

Como había indicado cuando presenté la propuesta de esta ponencia desde hace ya más de 10 años formo parte del grupo de redactores de exámenes de certificación de ruso en Cataluña.

Tenemos un sistema de trabajo muy bien organizado, unas especificaciones muy elaboradas y muy bien formuladas para poder redactar las pruebas de la mejor manera posible, unos criterios de evaluación basados en el currículo y contrastados con los objetivos marcados en el MCER. Recibimos una formación permanente y continuada como redactores de pruebas (no en vano los redactores de Cataluña lo que más valoran es precisamente la

formación). Y tenemos la suerte (a pesar de la desgracia de las estadísticas) de poder realizar un análisis casi individualizado de cada grupo de alumnos y de cada ítem.

Pero somos muy pocos profesores en la comunidad, todo queda en familia. Y eso quiere decir que tenemos la ventaja de seguir muy de cerca todo el proceso: elaboramos nuestros exámenes, los pilotamos, analizamos el pilotaje, redactamos la versión definitiva, examinamos y después analizamos los resultados. Porque, claro está, también somos profesoras y estamos por eso muy presentes en el proceso de enseñanza-aprendizaje.

Esta situación nuestra tiene sus ventajas y sus inconvenientes (ya que hoy hablamos de ventajas e inconvenientes...). Somos bastante conscientes de que **no tenemos la posibilidad de ver nuestro trabajo de lejos, a una cierta distancia. Y esa distancia sería buena, hasta necesaria.** Como dijo un poeta ruso, las cosas grandes se ven a distancia. Así que existe el peligro de perdernos en los detalles y no saber ver el conjunto de nuestro trabajo de forma adecuada.

Por otro lado **la gran ventaja** es precisamente poder ver los detalles y aprender a no perdernos en ellos. Eso nos exige siempre estar en forma, cuestionando lo que hacemos y cómo lo hacemos, no quedarnos satisfechas cuando un ítem da muy buen resultado en el pilotaje porque sabemos que nuestras estadísticas son relativas. Ni tampoco desesperarnos cuando un ítem (tan querido y tan trabajado) tiene que ser descartado después del pilotaje debido a la discriminación negativa.

Somos participantes muy activos en cada momento durante todo el proceso. Y en las escuelas más pequeñas a menudo examinamos a nuestros alumnos, es decir, tenemos que cambiar de papel y verlos no como alumnos sino como candidatos. Resulta difícil olvidarnos de que todo lo que saben o no saben es también el resultado de nuestro propio trabajo como profesoras.

En esta situación tenemos que ser muy profesionales y lograr el grado de objetividad más alto posible. En la situación de examen ya no son nuestros alumnos, son candidatos que tienen que demostrar que poseen el nivel de conocimiento del certificado que pretenden obtener.

Bajar al mínimo la subjetividad de la evaluación de los candidatos es el reto principal que tenemos que estar logrando continuamente como redactoras.

LA GRAMÁTICA ES UNA DULCE CANCIÓN

Esa doble personalidad nos ofrece muchas ventajas. Y nos plantea bastantes preguntas. Y muchas de esas preguntas están relacionadas con la parte de uso de la lengua.

- Buscar **equilibrio entre la competencia lingüística y la competencia comunicativa.** Hoy en día sabemos todos, y no lo discutimos ya, que la segunda es más importante, es el objetivo principal de la mayoría de nuestros alumnos, es lo que reclama el tiempo. Lo

demuestran también los resultados de un concurso que organizó el departamento de inglés en nuestra escuela: ha ganado la persona que había demostrado ser mejor comunicador aunque al nivel de competencia lingüística no era de los mejores, ni mucho menos.

- El problema es que a causa de la misma estructura de la lengua, **en ruso es muy importante tener muy buen conocimiento de gramática para lograr una comunicación exitosa** (se declina, las formas tienen un significado u otro, el cambio de forma podría romper la comunicación; lo mismo pasa con el sistema verbal y el significado de las formas aspectuales – me aconsejas o me lo ordenas, me dices que eso está prohibido o que es imposible de hacer –, todo eso depende de la forma utilizada y si no es la que tocaba la comunicación no tendrá mucho éxito). La gramática cada vez tiene menos peso dentro de la competencia comunicativa ya que se interpreta como un instrumento. Pero hay que tener en cuenta las particularidades de cada lengua. En ruso existe el concepto de la gramática comunicativa. El uso incorrecto de formas gramaticales puede romper la comunicación: uso de formas aspectuales, verbos de movimiento. Los alumnos siempre preguntan: ¿qué pasará si utilizo una forma en lugar de la otra?, ¿me entenderán los rusos? Creo que hoy en día ya podemos afirmar que nos hemos alejado del extremo de pensar que se puede hablar bien una lengua sin tener una sólida base gramatical.
- Por eso **insistimos en la necesidad de evaluar la parte de gramática y vocabulario**. Evidentemente en esta parte se evalúa más la competencia lingüística. Y la comunicativa se demostrará en las partes de producción (expresión e interacción escrita - EIE, expresión e interacción oral - EIO). Pero en nuestra lengua resulta muy importante esa parte porque **es muy difícil sin ella distinguir entre los errores de mayor o menor importancia desde el punto de vista comunicativo**.
- Otro aspecto que me gustaría mencionar en relación con esa parte son **los criterios**. Consideramos que para nuestra lengua decir que el texto (escrito u oral) tiene muchos errores de gramática, que es casi sin errores, que contiene errores propios del nivel, etc. es un criterio muy subjetivo, **al final será cada examinador que tendrá que decidir la importancia de cada uno de esos errores porque no tenemos una escala de errores más o menos importantes desde el punto de vista de la comunicación**. Siguiendo el principio de que cada tarea en un examen tiene que servir para evaluar una habilidad concreta, o un conocimiento concreto, muchas veces acabamos evaluando la competencia comunicativa en las partes productivas del examen por encima de la competencia lingüística. Y después nos “sorprenden” los resultados de algún que otro alumno que no ha aprobado la parte de uso de la lengua y en cambio tiene una nota bastante decente en la parte de EIO.
- **Sería interesante crear una lista de los errores gramaticales más importantes**. En Rusia no utilizan textos en esa parte. En cambio, los ítems (que también son cerrados) están ordenados y siempre se puede saber en qué parte se ha equivocado el candidato (formas de casos y preposiciones, formas verbales, formas aspectuales, estructura de oración compuesta). Esa posibilidad no la dan los textos o no siempre la dan, aunque evidentemente buscamos textos que tengan los elementos necesarios. Pero no siempre podemos retocar los textos escogidos y entonces surge la duda: **¿los candidatos se equivocan por no tener el nivel suficiente de competencia lingüística o por no**

entender el texto? En la fase de control seguramente deberíamos invertir el método: del textual (típico para el proceso de enseñanza-aprendizaje) al método funcional, que identifica las características sistemáticas de la gramática.

- Asimismo podríamos **introducir los ítems de relacionar en el uso de la lengua, sobre todo al NA** (nivel avanzado), donde estaría bien comprobar hasta qué punto los candidatos son capaces de interpretar, por ejemplo, el significado de una frase compuesta. Pensamos que esto podría aumentar la validez de la prueba, ya que en el NA es interesante evaluar si el candidato sabe utilizar verbos con prefijos diferentes (que cambian el significado), conoce el significado de oraciones simples y compuestas y su correlación, domina características estilísticas de las frases. Todo esto resulta difícil de comprobar utilizando solo los ítems cerrados de elección múltiple.
- **La parte de uso de la lengua representa en forma concentrada el contenido necesario para el nivel de la competencia lingüística.** Evidentemente, hay que tener en cuenta **qué** queremos evaluar, si todo realmente **merece ser evaluado** y **qué volumen es suficiente y necesario** para evaluar lo que queremos. B1 tendría que ser la base necesaria, pero la manera en la que se disponen los ítems ahora (dentro del texto en un orden mezclado) no nos permite ver qué exactamente no sabe el candidato, y le permitimos pasar al NA sin tener la base de NB (nivel básico) prácticamente.
- Resumiendo, la pregunta que nos hacemos es **si valdría la pena establecer un sistema más flexible de elaboración de los ítems** en la parte de uso de la lengua. Quizas, en ruso sería más interesante **tener la parte de ítems independientes más grande** de la que tenemos ahora (¿la mitad de los ítems como mínimo?).
- Estaría bien tener **un catálogo de errores.**

LA OBJETIVIDAD DE LO SUBJETIVO

En los exámenes tipo test se elimina el componente subjetivo que está bastante presente en las partes como EIE y EIO. **La evaluación de estas partes está basada en las impresiones del examinador.** Por eso es muy **importante especificar qué errores son poco significativos desde el punto de vista comunicativo y qué errores son muy significativos** desde este punto de vista. El criterio de “se entiende a primera lectura” o “el discurso es fácil de seguir” no dejan de ser subjetivos. Eso muchas veces da pie a discusiones en las sesiones de unificación de criterios.

No se puede convertir el control en la caza de errores. Los criterios de evaluación tienen que garantizar que se evalúa la competencia comunicativa ya que precisamente esta competencia hoy en día demuestra el dominio auténtico de la lengua. Y es lo que se les exige a nuestros alumnos en la vida real.

El tiempo es un componente importante porque sirve de indicador – hasta qué grado está formada la capacidad, hasta qué punto está desarrollada la competencia. De todas maneras **tenemos ciertas dudas sobre la cantidad de tiempo que tienen los**

candidatos para escribir dos redacciones. **O también sobre la cantidad de palabras** en la prueba de la EIE. Por ejemplo, en el NA se podría evaluar con 150 palabras sin ningún perjuicio, con este volumen ya se ve la capacidad de expresarse por escrito.

Existe un porcentaje bastante alto de “adivinar” la respuesta o poner la cruz por azar. Aquí el volumen también influye porque los candidatos empiezan a poner las cruces por azar. Eso lo comentan los alumnos.

La actualidad nos exige escoger no cualquier forma de control sino la forma óptima, la más válida y la más fiable.

Hoy en día el problema de la puntuación y del examen en la forma de test sigue siendo muy actual. Pero aún más actual es el problema de definición de los componentes de la competencia comunicativa a evaluar, así como de cuál es la mejor manera de evaluar esos componentes.

Las formas de control, como bien sabemos, vienen determinadas por lo que queremos evaluar. La pregunta qué, cómo y por qué evaluamos en nuestras pruebas es muy importante y siempre deberíamos tener la presente durante el proceso de elaboración de esas pruebas.

APRUEBAN LOS QUE TIENEN QUE APROBAR... / ? / !

Para garantizar la validez de los ítems tenemos que **estar seguros de que contienen solo una dificultad.**

Pero uno de los **inconvenientes de la prueba tipo test es que el candidato a menudo solo tiene que demostrar que ha reconocido lo que se pretende que se evalúe.** Eso no significa que será capaz de producir un discurso escrito u oral usando el mismo elemento lingüístico. Aquí podríamos encontrar una de las explicaciones de por qué los alumnos que pensamos que tienen que aprobar no aprueban. Y otros que no deberían aprobar pasan las pruebas con una puntuación bastante alta simplemente porque saben reconocer lo que se les pide que reconozcan. ¿Cómo solucionar esto? ¿Cómo subir el grado de fiabilidad de nuestras pruebas en este sentido?

El problema que tenemos hoy es precisamente el problema de evaluación de competencia real porque **resulta bastante difícil redactar unos criterios altamente objetivos y que, a ser posible, tengan en cuenta tanto la puntuación cuantitativa como la puntuación calificativa.**

Otro dato a comentar es que ha cambiado el perfil de nuestros candidatos. A causa de eso los parámetros como la intención comunicativa y el estatus social de los usuarios tienen

que estar presentes en las pruebas de lenguas extranjeras. ¿Por qué les hacemos escribir uno u otro texto?, ¿lo necesitan realmente en su vida real?

En cambio, la estructura combinada de nuestros test sí que resulta bastante válida porque se aproxima mucho a las condiciones reales de funcionamiento de una lengua (CE+EE; CE+EO)

El material para las pruebas de CE (comprensión escrita) y CO (comprensión oral) tiene que ser interesante para los candidatos, adecuado para su edad y posición social. ¿Seguro que lo tenemos contemplado en las especificaciones? Sí que tienen que estar presentes todos los estilos y registros. ¿Pero hace falta que escriban una cosa que nunca escribirían ni en su lengua materna?

¿CONTROLAMOS LO QUE QUEREMOS CONTROLAR?

La calidad del discurso es lo que adquiere mayor importancia. Por eso será necesario distinguir entre errores a nivel de competencia (lo que el candidato no sabe utilizar de acuerdo con las normas de la lengua) y errores a nivel de uso (cuando el candidato por alguna razón no llega a realizar su competencia de forma adecuada).

La evaluación de **la competencia comunicativa en cada nivel tiene sus particularidades** porque la competencia se vuelve cada vez más compleja. Entonces las formas de tareas de control deben cambiarse. Lo que, a su vez, conlleva cambios y correcciones en los criterios para la puntuación. Lo único que permanece sin cambiar es el concepto de que el control tiene que ser totalmente objetivo.

Y por lo que a **las habilidades productivas** se refiere, habrá que destacar que lo que nos exige la actualidad es **aprender a evaluar todos los componentes de la competencia comunicativa**: la lingüística (que esa ya la tenemos bastante controlada), la sociocultural y la pragmática.

Con estos últimos componentes tenemos todavía muchos problemas a la hora de evaluarlos porque no los tenemos ni tan controlados, ni tenemos elaboradas unas tareas que permitan evaluarlos.

Por ejemplo, en **el NA** donde los candidatos tienen que saber demostrar ya **el dominio de diferentes estilos y registros** no hay tareas para controlar si saben participar en una conversación oficial (como entrevista laboral, por ejemplo). Tenemos un modelo de examen de la EIO para el NA muy bien elaborado. Pero ¿nos permite saber si los candidatos pueden participar en una discusión improvisada, no preparada? Si hasta les ofrecemos una lista de preguntas para mantener la discusión. ¿No se supone que tendrían que saber hacerlo al NA por ellos mismos? En la vida real no tendrán listas de preguntas para hablar.

Un monólogo no preparado de carácter productivo (no reproductivo a base de un texto leído) tampoco lo tenemos en el NA. Y sería necesario tenerlo porque **aquí precisamente se podría evaluar la elección de elementos lingüísticos necesarios, el contenido del discurso, la coherencia y la cohesión del discurso, el grado en el que logran el objetivo comunicativo, la capacidad de autocontrol.**

Y en el **NI** sería importante controlar también **la capacidad de participar en una conversación de carácter habitual** como, por ejemplo, la conversación con el camarero en un restaurante. Eso es lo que la actualidad exige de nuestros alumnos y lo que muchas veces no saben hacer. Describen las fotos pero hace poco descubrí que no saben hablar con un camarero y lo tratan de tú ¡que en Rusia es totalmente inaceptable!

En Rusia se evalúa la capacidad del candidato de participar en una conversación, por ejemplo, en una oficina, en una consulta médica, con una amiga. Es decir, el componente sociocultural es muy presente en la evaluación. **Nosotros trabajamos todas esas competencias en clase pero no se ven reflejadas en la evaluación final.** Y resultan competencias más importantes que la capacidad de describir una foto.

Uno de los problemas o, mejor dicho, cuestiones pendientes es cómo evaluar la competencia pragmática. Hasta en los cursos superiores a los alumnos les cuesta muchísimo distinguir los matices del estado de ánimo de los personajes de una película, por ejemplo, o las intenciones de un personaje literario. Entienden las palabras pero siguen sin ver lo que hay detrás de esas palabras. Todos esos elementos con los que trabajamos en clase en el NA prácticamente no se ven reflejados en nuestros exámenes porque, hay que reconocerlo, **resulta muy difícil establecer los parámetros para evaluarlos.** Aquí también tenemos un camino que recorrer.

El significado y la importancia de elementos de introducción tendrían que ser evaluados sobre todo en el NA. Si no en la parte de la EO pues en la de uso de la lengua. Pero no se puede dejarlos fuera porque su semántica es uno de los objetivos marcados en el currículo del NA. Se podría evaluar los elementos que marcan la cortesía, sirven para llamar la atención, para estimular la respuesta, etc.

Otra cosa que también me gustaría comentar en relación con los exámenes de la EIO es que hasta ahora muchos colegas consideran que los candidatos tienen que transmitir de la mejor manera posible el contenido del texto leído. Es legítimo. ¿Pero realmente nos interesa eso a la hora de evaluar la capacidad de expresarse? Aquí hay dos dudas importantes. Primero **¿evaluamos la EO o la CE?** Si algún candidato no ha acabado de entender el texto que le ha tocado, **¿cómo evaluar la EO si no ha entendido el texto o no sabe qué decir sobre el tema?**

Los temas correspondientes a los intereses de los candidatos siempre favorecen más al cumplimiento satisfactorio de la tarea. ¡Los temas sobre los que no pueden hablar candidatos más jóvenes (el paro, el mundo laboral)! Pero, evidentemente, tiene que haber límites de trato personal porque los candidatos deben saber hablar sobre temas que no les

interesan pero que forman parte del currículo. Las exigencias sobre la competencia lingüística tienen que ser iguales para todos.

Las tareas de la EO tienen que formularse de manera más exacta posible para no dejar lugar a dudas e interpretaciones. La pregunta que a menudo hacen los alumnos - ¿qué pasa si lo digo así, si lo interpreto de esa manera?

El problema de los parámetros – por qué errores bajar más (los que tienen carácter comunicativo) o menos (los que no afectan la comunicación o no son tan significativos) la nota. Con esto también se podría relacionar el problema de **errores correspondientes a niveles más bajos** y que todavía se ven en candidatos de NA. ¿Considerar esos errores como correspondientes a ese mismo nivel? ¿Bajar más la puntuación por esos errores que por los que corresponden al nivel? Es decir, no perdonarlos ya al nivel más alto, **introducir un sistema de puntuación de EO y EE dependiente de la gravedad de los errores desde el punto de vista comunicativo.**

La realidad nos dicta sus exigencias. **Gana el que mejor comunica, no el que tiene más competencia lingüística. El que mejor comunica a pesar de hablar con los errores que a los profesores nos parecen imperdonables.**

En relación con la CO, ¿cuántas veces hay que presentar la grabación? En las clases, de acuerdo con el sistema de trabajar con textos auditivos, lo que hacemos más a menudo es presentar la grabación 3 veces (comprensión global, comprensión más detallada, comprensión prácticamente completa). ¿Por qué en el examen son solo dos? Y hay comunidades que sí las presentan 3 veces. ¿No deberíamos unificarlo?

El éxito o fracaso de un candidato depende también de su cultura general, de sus intereses, del sistema educativo del que viene y muchos otros elementos extralingüísticos. Creo que **tenemos abierta la cuestión de la diferencia de edades de nuestros candidatos** y diferentes grupos sociales.

En un examen los alumnos se bloquean si no entienden todas y cada una de las palabras. Eso solo significa que no saben utilizar las estrategias necesarias. En un examen tiene que haber palabras desconocidas (no en porcentaje muy alto evidentemente)

Resumiendo, la pregunta de la que nunca tenemos que olvidarnos y la que siempre permanece es **¿qué evaluamos con nuestras pruebas?: ¿La capacidad de aprobar el examen? ¿O las competencias?** ¿Podemos estar totalmente seguros de que nuestras pruebas garantizan el control de las competencias? **Porque si no lo podemos garantizar ganan los candidatos que “se han entrenado” para aprobar sabiendo lo que les exigirán, lo que se espera de ellos.**

LOS NATIVOS TAMBIÉN SE EQUIVOCAN

Un comentario más que me gustaría hacer. Y vuelvo otra vez a lo privilegiados que somos porque podemos ver y fijarnos en muchos detalles de nuestro trabajo. Siempre tenemos, sobre todo en el NA, **candidatos libres que son nativos** y vienen a hacer el examen para tener el certificado. Evidentemente nuestras pruebas no están pensadas para los nativos. Y por eso ellos aprueban sin mucho esfuerzo todas y cada una de las partes. Pero se puede observar una cosa que resulta significativa para nosotros. Llevo ya unos años apuntando en **qué ítems fallan los nativos** (que en principio no tendrían que fallar). Por ejemplo, en la parte de gramática no se equivocan, es lógico porque esa parte está pensada para controlar las dificultades y la competencia lingüística de los extranjeros. Por otro lado se equivocan a veces (no muy a menudo por suerte para nosotros) en la CO y CE. La pregunta es ¿por qué? Parece que se equivocan más a menudo en la CO: los nativos tampoco lo oyen todo y tampoco tienen tiempo de reaccionar... Pero también debemos tener presente que los nativos tienen mayor capacidad de interpretación precisamente porque se trata de la lengua materna. Y la interpretación puede dar un margen de error. Para nosotros eso resulta preocupante porque ponemos en el examen para nuestros alumnos unos ítems que se prestan a interpretación por parte de los nativos y seguramente no tenemos derecho a hacer eso. Y eso pasa a pesar de que normalmente en la fase inicial de la elaboración de las pruebas siempre intentamos contrastar los ítems probando cómo funcionan con nativos.

PREGUNTAS

- Las pruebas del examen de certificación son las mismas que se realizan en el aula previamente. Pero a menudo dan mejor resultado. ¿Por qué? ¿Tiempo? ¿Tensión?
- El análisis de resultados se tendría que hacer a base de pilotaje de una cantidad muy grande de candidatos. Como en nuestro caso resulta imposible tenemos que fijarnos de nuestra intuición. **Así que, para mejorar en este sentido, nos gustaría saber si hay alguna posibilidad de pilotar nuestras pruebas.**
- ¿Las partes (todas o alguna) tienen que ser eliminatorias?
- Referente a los errores, nos tenemos que plantear muy en serio en relación con la competencia comunicativa ¿qué errores son permitidos? ¿cuántos pueden ser?, porque será muy difícil encontrar el límite.

Aunque no representamos a la mayoría creemos que las observaciones que ofrecemos y las aportaciones que podemos hacer podrían ser interesantes para todos precisamente porque nuestra situación nos permite fijarnos en algunos detalles que, igual, se escapan a las observaciones los otros.