

Diseño de una prueba de clasificación de español no estandarizada

Design of a Spanish no standardized placement test

David Sánchez-Jiménez

Associate Professor. NYC College of Technology, CUNY

dsanchezjimenez@citytech.cuny.edu



David Sánchez Jiménez se licenció en Filología Hispánica en la Universidad de Salamanca en 2001, donde también cursó el Máster en Enseñanza de Español como Lengua Extranjera (2006). Completó su doctorado en Lingüística Aplicada en la Universidad Antonio de Nebrija en 2014. En 2014 recibió la beca postdoctoral Andrew W. Mellon Postdoctoral Fellowship en Colgate University. Tras distintas estancias en instituciones internacionales (Philippine Normal University, University of the Philippines, Fazekas Mihály Gimnázium (Budapest) y University of Washington), en 2016 comenzó a trabajar en New York City College of Technology (CUNY), donde enseña en la actualidad

Resumen

El presente estudio propone una nueva prueba de clasificación en línea específicamente creada para los estudiantes de español de New York City College of Technology (NYCCT), una institución caracterizada por la elevada diversidad de su comunidad educativa. En 2017 el Departamento de Humanidades decidió crear su propia prueba de clasificación de lenguas. En el caso del idioma español, esta se creó y se pilotó en 2018 con una muestra de 144 alumnos distribuidos en 3 niveles: inicial, intermedio y avanzado alto. Los resultados de esta práctica mostraron datos relevantes que permitieron identificar factores inesperados en el análisis, como la falta de compromiso con la prueba, la gran cantidad de alumnos colocados fuera de su nivel y el alto número de hablantes de herencia con un elevado nivel de comunicación que concurrían a las clases de español como lengua extranjera no solo en el nivel avanzado, sino también en el intermedio.

Abstract

This study proposes a new online placement test specifically created for students of Spanish at New York City College of Technology (NYCCT) - an institution characterized by the great diversity of its educational community. In 2017, the Humanities Department decided to create its own foreign language placement test. In the case of the Spanish language, this was created and piloted in 2018 in a sample of 144 students distributed in three levels - basic, intermediate and high advanced. The

results of these tests showed relevant data that allowed to identify unexpected factors in the analysis, such as the lack of dedication to the test, the large number of students placed outside their level, and the high number of Spanish heritage speakers with good communicative skills who attended these classes not only in the advanced, but also in the intermediate level.

Palabras clave

Evaluación, pruebas de nivel, aprendizaje de la lengua, autoevaluación, competencia lingüística.

Keywords

Assessment, placement test, language learning, self-assessment, linguistic competence.

Introducción

La prueba de clasificación tiene como objetivo recopilar información referida a los fallos y aciertos producidos por el aprendiz al resolver una tarea en una determinada área de conocimiento, cuyo fin último es el de proporcionar información útil sobre el nivel o curso en el que el candidato debe inscribirse en el centro educativo en el que se cursa sus estudios. En este artículo se informa de una nueva prueba de clasificación de español en línea desarrollada en New York City College of Technology (NYCCT), una de las 25 instituciones de educación superior que conforman el consorcio de la Universidad de la Ciudad de Nueva York (CUNY).

Con más de 275.000 alumnos procedentes de 205 países diferentes, CUNY ha apostado desde sus inicios por la diversidad y la inclusión de estudiantes de diferentes razas, edades y clases sociales, aspirando –como compromiso primero y objetivo último en la misión de su empresa– al fomento de la movilidad social de los alumnos matriculados en sus aulas independientemente de su procedencia. Dentro de este sistema universitario, NYCCT recibe anualmente más de 17.000 estudiantes procedentes de más de 140 países diferentes (The City University of New York, 2019).

Esta singular idiosincrasia explica que los exámenes de clasificación comerciales que se han venido utilizando convencionalmente para medir el nivel de lengua en español de la población de NYCCT no resulten adecuados. Por este motivo, en 2017 el Departamento de Humanidades decidió crear su propio examen de clasificación de lenguas. En el caso del idioma español, este se creó y se pilotó en 2018 con una muestra de 144 alumnos distribuidos en 3 niveles: inicial, intermedio y avanzado alto. El análisis de los resultados obtenidos y las entrevistas realizadas posteriormente a los estudiantes que realizaron esta prueba permitieron identificar algunos aciertos en la administración de este examen, los cuales se comentarán a lo largo de este artículo. También se aludirá a las diversas facetas en las que esta tarea debe ser mejorada para ofrecer un mejor servicio a los estudiantes matriculados en el Programa de Español en NYCCT.

De modo general, el ejemplo de este caso servirá como modelo a otras instituciones que decidan adoptar una prueba local de clasificación en sus centros de lenguas. Esta propuesta parte de una iniciativa que pretende promover la creación de una prueba de clasificación adaptada a las necesidades específicas de la institución, lo cual resulta especialmente conveniente en el caso de programas de lenguas insertos en escuelas o universidades que a causa de la heterogeneidad de su alumnado no se encuentran alineados con las directrices prescritas en los modelos estandarizados de

medición de lenguas. La propia singularidad del programa de cada institución permitirá entender las decisiones que el centro deba tomar con respecto a la selección de la prueba de clasificación.

1. La evaluación y la prueba de clasificación

La evaluación es un aspecto esencial del diseño curricular de cualquier programa de lenguas. De acuerdo con la división establecida por Bordón (2015: 10), debe distinguirse entre los conceptos de *examinar* (medir la actuación del candidato) y *evaluar* (recoger y analizar datos significativos acerca de una situación en el aula para tomar decisiones a partir de ellos). Así, independientemente del tipo de prueba que se use, un examen constituye un instrumento para evaluar, es decir, para tomar decisiones académicas y administrativas más allá de la función penalizadora clásica con la que en ocasiones se asocia esta práctica, como refieren Bordón (1990), Yagüe (2010) y Fernández y Staire (2017). En cuanto a su tipología, Brown (1996: 1) diferencia entre 4 tipos distintos, agrupados en función de los datos que aportan sus resultados: los que permiten hacer decisiones a nivel del programa del curso (pruebas de *suficiencia* y de *clasificación*) y aquellos que asisten al instructor para introducir cambios en la clase (pruebas de *diagnóstico* y de *aprovechamiento*). En el caso que nos ocupa, y en consonancia con lo referido por este autor (Brown, 1996: 13), se postula que la prueba de clasificación de una institución debe ser diseñada para agrupar a los estudiantes en correspondencia directa con un programa específico de lenguas y resultar apropiada para el programa en cuestión, de modo que su utilidad sea manifiesta para todos los miembros de la comunidad educativa. En esta función reside una de las características esenciales que debe cumplir este examen, el cual se pone al servicio de las necesidades de aprendizaje de un grupo particular de alumnos en el contexto de uso concreto de la lengua que se está adquiriendo (Bachman y Palmer, 1996: 17). En palabras de Bordón (2006: 81):

La ventaja de diseñar exámenes de clasificación originales es que serán más coherentes con los niveles existentes en el propio centro y de esa manera, serán, también, más efectivos. El proceso de enseñanza-aprendizaje se beneficia, tanto a nivel individual como colectivo, si quienes integran el grupo de clase parten de unos conocimientos (o desconocimientos) similares y tienen unas necesidades comunicativas parecidas. (Bordón, 2006: 81)

En el caso concreto de NYCCT no existe un título de grado en la disciplina de español, por lo que las asignaturas cursadas en esta lengua forman parte de los requisitos que los estudiantes deben completar para finalizar su carrera universitaria, como ocurre igualmente en muchas otras universidades a lo largo y ancho de los Estados Unidos. La metodología de enseñanza seguida en el centro (ligada a las actividades estructurales propuestas en el libro de texto elegido por la coordinación del Programa de Lenguas Extranjeras) y la formación tradicional en la enseñanza de lenguas que comparte la mayoría del profesorado, como consecuencia, definen el contexto de uso que los estudiantes de NYCCT hacen de la lengua. Esto debe ser tenido en cuenta al diseñar el instrumento creado para agrupar a los estudiantes en cursos, que además debe corresponderse con la metodología de enseñanza utilizada en la institución y el tipo de actividades desarrolladas en el aula, así como con la capacidad para aportar información relevante sobre el proceso de evaluación tanto a los docentes como a los estudiantes (Bordón, 2006: 85).

Nuestra prueba de clasificación fue creada con el propósito definido de distribuir a los estudiantes de NYCCT en los distintos niveles establecidos en el Programa de Español (véase § 2) según el conocimiento lingüístico que posee cada uno de los

individuos evaluados. Este debe coincidir, por lo tanto, con los objetivos y los contenidos delimitados para cada uno de los cursos descritos en el apartado 2. No obstante, con el fin de aumentar su fiabilidad, el análisis de estos aspectos lingüísticos se realizó mediante diversos métodos y tipos de actividades (autoevaluación, selección múltiple, producción escrita e interacción oral), como sugieren Alderson, Clapham y Wall (1995: 45).

En base a las características propias del centro, el equipo encargado de diseñar la prueba de clasificación tuvo que tomar una serie de decisiones drásticas y sistemáticas para adaptar el examen al Programa de Español existente, como la de prescindir de la evaluación de las habilidades lingüísticas individuales de los candidatos para comunicarse en las diferentes destrezas de la lengua, así como su competencia comunicativa en cada una de estas áreas. La prueba se concibió estrictamente como una herramienta pedagógica que facilitara la programación de los cursos en respuesta a las necesidades educativas concretas de los estudiantes de NYCCT, las cuales se detallarán en el apartado 3.

De este modo, contradiciendo la bibliografía más actualizada sobre la construcción de pruebas en segundas lenguas (Martín Leralta, 2009; Yagüe, 2010; Baralo, 2012; Ciarra Tejada, 2013; Ferreira, Vine y Elejalde, 2013; Fernández y Staire, 2017), se primó el uso de instrumentos empleados para medir el conocimiento gramatical y léxico. Con esta decisión se buscaba valorar el conocimiento adquirido durante los años de formación en la escuela primaria y secundaria, pero también en otras circunstancias diferentes de inmersión lingüística, ya que como bien señala Ortega (2019), las formas en que los hablantes adquieren una segunda lengua y su capacidad para comunicarse en ella no siempre responden a los patrones institucionales de instrucción adquiridos en contextos formales de aprendizaje de la lengua (especialmente los hablantes de herencia).

Aparte de las comentadas anteriormente, se tomaron otra serie de decisiones relacionadas con el diseño de la prueba a partir de la observación de los resultados obtenidos en los exámenes de clasificación utilizados previamente en nuestra institución. Considerando estas experiencias, se tuvieron en cuenta un conjunto de características específicas en la construcción de la prueba, las cuales se exponen en el apartado 3. En la siguiente sección se explica el funcionamiento del programa de lenguas en NYCCT, ya que su constitución idiosincrásica justifica la modalidad de la prueba de clasificación descrita en este artículo.

2. La enseñanza de lenguas en NYCCT

En la actualidad, el Programa de Lenguas Extranjeras del Departamento de Humanidades de NYCCT ofrece el estudio de 4 lenguas extranjeras (español, árabe, chino y francés) y cuenta con 3 profesores titulares y 14 adjuntos para impartir las asignaturas correspondientes de este programa. Cursar al menos 9 créditos en una de estas lenguas es un requisito para todos aquellos alumnos que quieran graduarse en la diplomatura de *Liberal Arts* (Artes Liberales) de la Facultad de Artes y Ciencias, una de las más demandadas por parte de los estudiantes que desean proseguir sus estudios en otras disciplinas relacionadas con las Artes y las Ciencias en NYCCT. Para el resto de estudiantes, el estudio de una lengua extranjera supone una asignatura electiva con la que pueden cumplimentar 3 créditos universitarios.

En total, más de 900 estudiantes cursan asignaturas relacionadas con el estudio de una lengua extranjera en NYCCT cada año. De entre ellas, el español es, sin duda, la lengua más popular y más estudiada de las que se ofrecen en NYCCT debido a su extensa presencia y a la gran importancia socio-económica que esta posee en la ciudad

de Nueva York (véase Otheguy y Zentella, 2012). Los datos recogidos en la última década por el Comité de Lenguas Extranjeras en el Departamento de Humanidades en NYCCT, apreciables en el Gráfico 1, así lo corroboran:

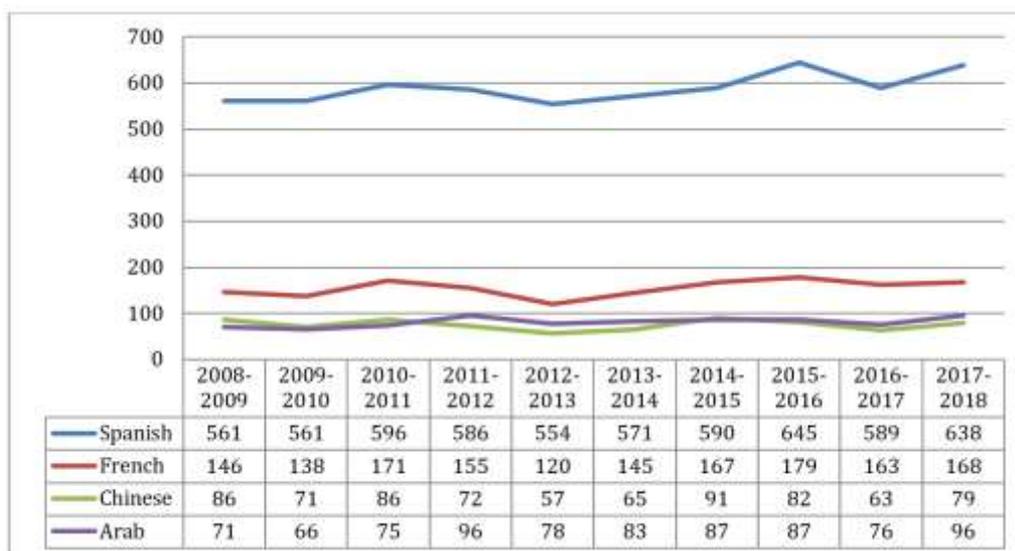


Gráfico 1. Datos de matriculación histórica en el Programa de Lenguas Extranjeras de NYCCT (2008-2018).

Fuente: elaboración propia

La matriculación de los estudiantes en el Programa de Español en NYCCT alcanza cada semestre los 250-300 candidatos (aproximadamente el 65% del total de los estudiantes matriculados en el Programa de Lenguas Extranjeras), que se distribuyen en dos secuencias conformadas por 9 cursos, de los cuales se ofrecen 15 secciones diferentes cada semestre con un máximo de 25 estudiantes por clase (Aire, 2019).

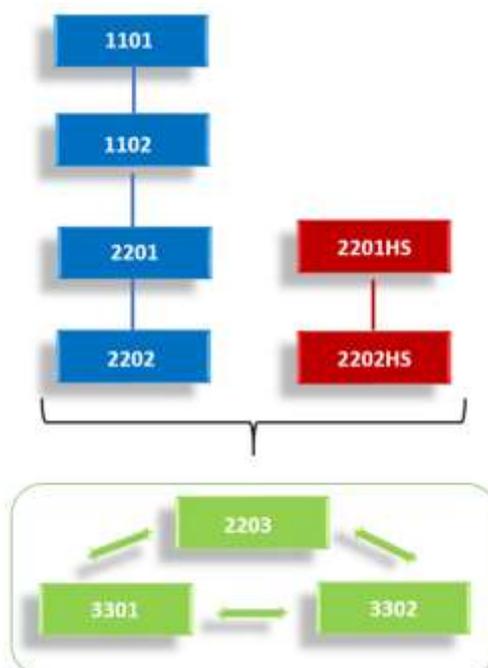


Figura 1. Cursos de español ofrecidos en el Departamento de Humanidades de NYCCT.

Fuente: elaboración propia

Como se aprecia en la Figura 1, la primera dificultad para la clasificación de los estudiantes en el Programa de Español estriba en la necesidad de identificar la naturaleza de los estudiantes de español como segunda lengua o como lengua de herencia (cursos marcados con el sufijo HS). Esta es una diferenciación fundamental para asegurar el éxito del programa, ya que los estudiantes de español como lengua de herencia demandan una metodología específica de enseñanza (Valdés, 1997; Beaudrie, Ducar y Potowski, 2014).

En la primera secuencia existen 4 niveles delimitados por la estructura del libro de texto utilizado como base para confeccionar el currículo de estos cursos, *Puntos de Partida*: SPA 1101 (capítulos 0-6), SPA 1102 (capítulos 7-12) y SPA 2201 (capítulos 13-18). A pesar de la descripción del material como una herramienta para conseguir una sólida base en el desarrollo del lenguaje comunicativo (Dorwick, Pérez-Gironés y Becher, 2016), la metodología utilizada en el diseño de las unidades y las actividades recuerda al enfoque nocional-funcional de la enseñanza de la lengua, caracterizada por una fuerte tendencia a la fijación de estructuras gramaticales por repetición y extensos listados de vocabulario agrupados semánticamente en cada capítulo. En el nivel SPA 2202 se hace un repaso general de los temas de gramática con el manual *En Contacto: Gramática en Acción* (Gill, Wegmann y Mendez-Faith, 2011), de estructura similar al anterior, y se insiste en la destreza de la expresión escrita con el fin de que los estudiantes lleguen a producir textos expositivos y argumentativos en español al final del curso.

La segunda secuencia se compone únicamente de dos cursos (SPA 2201HS y SPA 2202HS) dedicados a la práctica y el aprendizaje de la lengua de hablantes bilingües que tienen el español como lengua materna. En el primero de ellos se refuerzan las destrezas receptivas y se introducen temas socioculturales de interés para la comunidad latina en los Estados Unidos (inmigración, educación bilingüe, trabajos que exigen el dominio del español en Estados Unidos, la situación del español de este país y sus características morfosintácticas, etc.). En su continuación, SPA 2202HS, los objetivos se centran en la producción escrita de textos expositivos y argumentativos en diferentes géneros académicos, la enseñanza de recursos discursivos y de estrategias para mejorar la exposición oral y la escritura académica, la explicación de las estructuras sintácticas complejas del español y la presentación de temas de investigación relacionados con la historia y la cultura de los países en los que el español es lengua oficial.

A los dos últimos cursos de literatura (SPA 3301 y SPA 3302) se puede acceder al término de ambas secuencias. SPA 2203 es un curso de contenido para fines profesionales médicos impartido en español.

La razón por la cual nos propusimos modificar la prueba de clasificación existente surgió a partir de la detección por parte del Comité de Lenguas Extranjeras en 2016 de una serie de irregularidades asociadas al nivel lingüístico que presentaban los alumnos matriculados en los cursos de español. Entre estos figuraban la existencia constante de diferentes niveles de dominio de la lengua mezclados en los cursos y la presencia de estudiantes de español como lengua de herencia en los niveles iniciales. La observación de estos fenómenos repetidos de forma sistemática en el Programa de Español fue determinante para iniciar una investigación-acción cuyo objetivo principal radicaba en equilibrar los niveles de dominio de la lengua en cada curso a través del diseño de una nueva prueba de clasificación y la creación de nuevos cursos en el Programa de Español que dieran respuesta a las necesidades educativas que presentaba la población estudiantil de NYCCT (SPA 2203 fue creado en 2017 y SPA 2202HS en 2018).

3. Diseño de la prueba

3.1. Historia del examen de clasificación en NYCCT

Durante las décadas en las que los cursos de español han venido ofreciéndose de forma regular en NYCCT, la diversa procedencia de sus estudiantes y de los contextos en los que adquirieron sus habilidades lingüísticas en lengua española han contribuido a la elevada heterogeneidad de niveles mezclados en los cursos de español ofrecidos en nuestra institución. Como referían Fairclough y Ramírez (2009) al diseñar su prueba de clasificación, una de las mayores dificultades a las que se enfrenta el autor al crear este tipo de exámenes en los Estados Unidos radica en la heterogeneidad de la población diversa y multilingüe que ingresa cada año en los programas universitarios en este país. Estos estudiantes –nacionales e internacionales– provienen de distintas escuelas que cuentan con diferentes programas educativos en otros estados y países, y además poseen muy diversos perfiles socioeconómicos y lingüísticos. Sin embargo, estos no han sido los únicos factores que han incidido en tal situación, pues los diferentes intentos por clasificar a los estudiantes en los correspondientes grupos siempre se han realizado con herramientas inadecuadas y bajo parámetros que no se correspondían con las tareas y las tipologías de actividades desarrolladas a diario en el aula.

En los orígenes del Programa de Español, los alumnos de NYCCT realizaban una prueba corta en papel compuesta por 30 ítems de selección múltiple con 3 opciones. Estos eran después corregidos con una plantilla por la secretaria del departamento (hablante monolingüe de inglés), quien se encargaba de comunicar a los candidatos su puntuación y de asignarles un curso basándose en los resultados obtenidos en la prueba. Desoyendo la voz de los expertos con respecto a la conveniencia de desarrollar una prueba de clasificación local acorde con las necesidades pedagógicas de la institución (Alderson, Clapham y Wall, 1995; Bachman y Palmer, 1996; Brown, 1996; Bordón, 2006, 2015; Potowski, Parada y Morgan-Short, 2006; Fairclough y Ramírez, 2009; Martín Leralta, 2009; Baralo, 2012; Ferreira, Vine y Elejalde, 2013; Pérez Guillot, Zabala y Jaime, 2013; Johnson y Riazi, 2017), el Departamento de Humanidades comenzó a utilizar el examen comercial estandarizado *Avant Assessment* (s.f.) para español, chino, árabe y francés. Esta, sin embargo, no fue una decisión aislada tomada por nuestra institución. Como Alderson, Clapham y Wall (1995: 241) observan, es generalizada la dependencia que existe en Estados Unidos sobre el uso de este tipo de pruebas estandarizadas. En lo que se refiere a la versión que nosotros manejamos en español, no tardamos en identificar ciertas inconveniencias que contravenían el uso prolongado de este, por otra parte, costoso procedimiento de análisis del nivel de lengua, las cuales se comentan a continuación.

En primer lugar, los estudiantes que declaraban –en la primera parte de autoevaluación de la prueba– no haber tenido ningún contacto con la lengua a lo largo de su vida, estaban exentos de tomar el examen y eran automáticamente ubicados en el curso elemental nivel 1 (SPA 1101). Muchos alumnos utilizaban esta opción para acceder a un curso de lengua de nivel más bajo del que les correspondía con la finalidad de obtener una calificación más alta sin realizar un alto esfuerzo en el curso asignado. Además de esto, la estructura misma de la prueba resultaba ciertamente desequilibrada. El mayor porcentaje del examen lo conformaban preguntas de comprensión lectora graduadas en progresión ascendente de longitud y dificultad en un formato que ciertamente resultaba atractivo. Las preguntas sobre estos textos mantenían la misma dinámica que las tareas de gramática y vocabulario: selección múltiple sobre cuatro opciones con tres descriptores.

En contraste, en el apartado de las destrezas productivas, la prueba se limitaba a ofrecer un contexto situacional en el que el candidato debía interactuar con la actividad aportando su opinión personal en respuesta. Había una sola pregunta breve de expresión oral y otra de escritura a las que los estudiantes respondían frecuentemente con una oración simple y, en ocasiones, con una única frase de sintaxis atrofiada, la cual era calificada de forma aleatoria por un evaluador externo. Por otra parte, el micrófono para grabar la respuesta a la situación oral solía no funcionar o de manera sistemática no era compatible con el *software*, hecho que desvirtuaba la media de la puntuación final entre las distintas destrezas analizadas.

Por último, los administradores de la prueba insistían en mantener la longitud original del examen por cuestiones de fiabilidad, el cual estaba programado para unos 80 minutos de duración y trufado de actividades redundantes de comprensión lectora, con temática distinta pero formato similar. Sin embargo, comprobamos con sorpresa que el tiempo total de realización de la prueba variaba desde los 10 o 15 minutos hasta picos de 180 minutos.

Motivados por los datos obtenidos, decidimos crear una segunda parte de verificación de los resultados de la prueba a través de una breve entrevista post-examen. Al preguntar a los estudiantes a qué se debían estas bruscas desviaciones sobre el tiempo estimado para completar la prueba (80 minutos), estos respondían que las actividades propuestas en *Avant Assessment* (s.f.) les parecían tediosas y repetitivas, por lo que decidían no completarlas. En aquellos casos en los que el tiempo se extendía a los 180 minutos, los alumnos comentaron que habían estado trabajando en otras tareas en la computadora al mismo tiempo que realizaban la prueba, las cuales iban alternando de manera asistemática y desinteresada.

Es obvio que la atención de un estudiante que tomaba el examen mientras resolvía sus tareas de matemáticas o consultaba *Facebook*, no se correspondía con el escenario más adecuado para obtener datos válidos sobre su conocimiento lingüístico. De este modo, comprobamos mediante la entrevista que los datos que arrojaba *Avant Assessment* (s.f.) no siempre se correspondían con el verdadero perfil lingüístico del estudiante.

Es posible que esta misma prueba estandarizada esté funcionando de manera adecuada en otras universidades, no lo ponemos en duda. Lo cierto es que en nuestra institución no se adaptaba a las necesidades específicas que presentan nuestros estudiantes: jóvenes neoyorquinos con obligaciones laborales a tiempo completo o adultos con dependientes a su cargo que tratan de completar sus estudios universitarios con el propósito de conseguir una promoción en su trabajo o un mejor empleo. Todos ellos estudian una carrera aplicada relacionada con una profesión o son estudiantes de carreras relacionadas con la tecnología aplicada. Su interés y motivación real al estudiar los cursos de lengua es relativa y nunca prioritaria en su carrera universitaria, en ocasiones suscitada por un sentimiento emocional de identidad en el mejor de los casos, que suele ser el de los hablantes de herencia de español (normalmente inmigrantes de segunda o tercera generación) o el de aquellos estudiantes que tienen la oportunidad de utilizar la lengua en su entorno familiar o profesional.

3.2. Propuesta de la nueva prueba de clasificación

Como reacción a las experiencias previas comentadas en el subapartado anterior, decidimos diseñar nuestra propia prueba de clasificación en papel. Para ello, se incluyeron preguntas de gramática y vocabulario referidas a los temas estudiados en los cursos de español en NYCCT con el formato de ejercicios de selección múltiple, y se incorporaron 8 preguntas de escritura (relacionadas con las situaciones planteadas en

diversas fotografías o dibujos) y un examen oral de verificación con uno de los profesores titulares de la lengua a evaluar.

Con este nuevo modelo se respondía a uno de los mayores inconvenientes observados en la prueba anterior, la duración total del examen. A pesar de ello, pronto se identificaron nuevos problemas: el tamaño de la letra elegida dificultaba la legibilidad del texto, algunos de los exámenes en papel sufrieron daños o deterioros, era difícil almacenarlos en la oficina o hallar con rapidez la prueba correspondiente al candidato que se presentaba con posterioridad a la entrevista oral. El mismo hecho de revisar la prueba y calificarla en su presencia demoraba el proceso. Al mismo tiempo, comprobamos que en ocasiones el candidato dejaba sin responder algunas preguntas en el examen escrito, quizás por despiste o tal vez por el desinterés con el que realizaba la tarea. Por todos estos motivos, el paso lógico siguiente exigía agilizar la administración de la prueba con una versión remota computarizada en un formato más atractivo para el estudiante, que se corrigiera de manera automática, le obligara a incluir una posible respuesta para cada una de las preguntas y que además ofreciera los resultados con datos agregados para su fácil manejo.

De este modo, se resolvió alojar la prueba de clasificación en el *software* en línea de *Google Forms* para mayor comodidad y accesibilidad de los estudiantes. Mediante *Google Forms* se pudieron crear cuestionarios, ejercicios de selección múltiple, recuperar textos escritos de longitud considerable, insertar fotos y realizar, además, una serie de funciones estadísticas que no solamente permitían agrupar los datos de los estudiantes, sino también analizar el funcionamiento del propio instrumento de manera totalmente gratuita. Un ejemplo de este uso es la comprobación de aciertos por cada número de ítem, información mediante la que se puede conocer mejor cuáles son los ítems que discriminan más y menos en cada una de las secciones establecidas (inicial, intermedio y avanzado alto).

Estas modificaciones nos permitieron revisar el diseño y la estructura de la prueba de clasificación administrada previamente en papel, incorporándose a la prueba las manifiestas ventajas que diversos creadores de exámenes computarizados habían puesto de relieve con anterioridad (García Labora y Navarro, 2008; Pérez Guillot, Zabala y Jaime, 2013; Kim et al., 2018; Long et al., 2018): puntuación automática e inmediatez en el acceso a los resultados, optimización del tiempo empleado en la tarea y de los recursos humanos utilizados, disminución de los costes de administración de la prueba, facilidad al establecer un control de calidad y de manejo de los datos, flexibilidad en el tiempo empleado en administrar el examen y autonomía del lugar elegido para realizarlo, familiaridad con las nuevas tecnologías o la rebaja de los niveles de ansiedad que experimentan los candidatos (en contraste con quienes realizan esta tarea en papel en la institución).

El examen está disponible en línea y se puede acceder a él de forma gratuita en la página del Departamento de Humanidades de NYCCT. En los siguientes apartados se describen el formato, la estructura y las fases de elaboración de la prueba con más detalle, que pueden servir de modelo a proyectos similares en otras instituciones.

3.3. Formato y estructura

En los apartados anteriores se justificaron algunas de las decisiones tomadas para diseñar una prueba de clasificación que se adaptara a las condiciones específicas que presenta la institución NYCCT y al tipo de alumnado que accede a esta universidad. Como respuesta a la experiencia negativa obtenida con las pruebas realizadas anteriormente en el centro, se prefirió crear un examen local computarizado en línea que abarataba sustancialmente los costes de administración, redujera la duración del examen

—que repelía a los estudiantes en la práctica anterior— y que, en consecuencia, aumentara la eficacia en el desarrollo de la prueba.

Para crear este examen se tuvieron en cuenta los criterios descritos por Bachman y Palmer (1996: 60-64), especialmente los tres que de forma repetida coinciden en señalar los expertos en esta materia como centrales (Bordón, 1990, 2015; Alderson, Clapham y Wall, 1995; Bachman y Palmer, 1996; Brown, 1996):

Fiabilidad: la prueba resulta consistente en sus resultados y en la medición de los niveles establecidos. El pilotaje del examen proporcionó resultados similares para candidatos de diferentes grupos en el mismo nivel y diferentes para los estudiantes agrupados en distintos niveles (véase § 3.3. *fase de experimentación*). La administración de esta prueba en el semestre de la primavera de 2019 ha confirmado estos datos, en la que han resultado escasos los cambios de los estudiantes que no habían sido agrupados en el curso correspondiente a su nivel de lengua.

Viabilidad: el formato en línea mediante el *software* de *Google Forms* facilitó la administración del examen, mejorando la legibilidad de la versión en papel, la flexibilidad de tiempo y lugar en que se realiza la prueba y la amigabilidad del entorno informatizado. También se tradujeron las instrucciones al inglés para hacer comprensible el procedimiento de la prueba para aquellos candidatos con un menor dominio de la lengua española.

Validez: los expertos señalan que este asunto es el más complejo y el que requiere más atención para los creadores de exámenes de lenguas, pues está relacionado con el objeto mismo de la medición de la prueba. Nuestro examen, por ejemplo, prescinde de evaluar la competencia comunicativa que el candidato posee de la lengua para fijarse en los aspectos gramaticales y léxicos estudiados en el libro de texto de acuerdo con el programa de cada curso y nivel (véase § 1). Con ello, no pretendemos evaluar el conocimiento que el candidato posee de la lengua (capacidad para comprender, expresarse, producir o interactuar en situaciones de uso real de la lengua), sino el saber acumulado sobre los elementos constitutivos de la estructura de la lengua.

El análisis de estos contenidos lingüísticos se realiza mediante diversos métodos, lo que redundará en una mayor fiabilidad de la medición de la prueba (Alderson, Clapham y Wall, 1995: 45), y justifica su división en 4 grandes bloques: autoevaluación, gramática y vocabulario, producción escrita e interacción oral. Primeramente, la prueba se inicia con un apartado en el que se repasa la trayectoria académica y lingüística del estudiante en relación a su conocimiento de la lengua española, lo que nos permite obtener información de tipo cualitativo sobre el candidato: si sus familiares hablan español, si utilizan la lengua en un contexto profesional, viajes realizados y estancias en un país hispanohablante, educación parcial en uno de esos países, si tomó cursos en español de AP (*Advanced Placement*), etc.

Sobre esta práctica, diversos estudios se han encargado de analizar la fiabilidad de las respuestas obtenidas en la prueba de autoevaluación en los exámenes de lengua, entre ellos los de Edele et al. (2015) y Ferris, Evans y Kurzer (2017). Ambos coinciden en señalar que las diferencias entre los datos obtenidos en el examen de autoevaluación y el examen que evalúa los contenidos lingüísticos no son significativas, aunque todavía instan a los creadores de exámenes de lengua a utilizar las dos con el fin de reforzar la fiabilidad de la prueba. También observan estos autores que la autoevaluación resulta más efectiva en aquellos candidatos que tienen mayor competencia lingüística y experiencia estudiando lenguas.

En la segunda parte, meramente cuantitativa, se distribuyen 60 preguntas objetivas de selección múltiple en tres bloques con contenidos del nivel inicial (SPA 1101 y SPA 1102), intermedio (SPA 2201) y avanzado alto (SPA 2202). En la siguiente sección de la prueba se incorporan dos tareas subjetivas, una de escritura con 8 preguntas de desarrollo (sobre información personal, uso de los verbos de presente de indicativo e identificación de objetos cotidianos (SPA 1101), contraste de pasados usados en la narración (SPA 1102) y situaciones hipotéticas para usar los tiempos de futuro, condicional y oraciones compuestas de valor hipotético con subjuntivo (SPA 2201 + SPA 2202) y una tarea oral de 10-15 minutos de duración y de calificación holística, la cual presenta una estructura similar a la de la prueba escrita.

En esta última parte oral del examen se revisa la trayectoria académica descrita por el estudiante al comienzo del ejercicio, se contrastan rápidamente la puntuación de la prueba objetiva en cada una de sus partes con los resultados obtenidos en la escrita y se corroboran todos estos datos en una conversación con preguntas que buscan la precisión gramatical y léxica sobre contenidos específicos del sistema lingüístico (véase 3.4.). En el caso de los hablantes de español como lengua de herencia¹, se distingue entre aquellos que poseen un vocabulario receptivo (a los que se invita a estudiar en uno de los cursos de L2), los que pueden comunicarse de manera fluida en español con un vocabulario limitado, pero no han adquirido un grado de literacidad suficiente para comunicarse por escrito en un contexto académico (SPA 2201HS) y los que pueden leer y escribir sin dificultad, pero necesitan mejorar cuestiones relacionadas con la gramática, puntuación o aspectos discursivos orales y escritos (SPA 2202HS).

Una vez que se determina el curso específico del Programa de Español al que pertenece el candidato mediante el análisis contrastado de esta información, se le otorga permiso para matricularse mediante la introducción de su número de estudiante y el curso asignado en el sistema informatizado de la universidad *CUNYfirst*. No obstante, el proceso de clasificación cuenta todavía con un paso más de verificación del nivel de lengua. Al comienzo del semestre se habilita un periodo de tiempo –que no suele exceder las dos primeras semanas de clase– en las que el instructor que imparte el curso puede acomodar al alumno en el nivel que considera apropiado cuando determina que este no ha sido ubicado apropiadamente en su clase, previa comunicación al Comité de Lenguas Extranjeras para la revisión del caso.

¹ Uno de los resultados más positivos logrados en el proceso de creación de esta prueba de clasificación tiene que ver con la visibilidad otorgada a los estudiantes de español como lengua de herencia en NYCCT. Hasta este momento solo existía un curso (SPA 2201HS) –que no solía ofrecerse por el bajo número de estudiantes matriculados–, ya que predominaban los cursos mixtos en los que estos alumnos convivían mezclados con hablantes monolingües de inglés (o multilingües de inglés y otras lenguas) que estudiaban el español como una L2. A raíz de la identificación de este grupo y de la labor pedagógica realizada al explicar que se debía atender a sus necesidades educativas y perfil propios (adquisición de la lengua en contextos naturales, desarrollo típicamente elevado de los niveles de lengua oral cuando su literacidad y metaconocimiento lingüístico tiende a ser débil), se consiguió ofrecer el curso de SPA 2202HS en 2018, tras ser aprobado por el Comité Curricular y el Consejo de NYCCT. Esta situación es bastante frecuente en los Estados Unidos (para una revisión del tema, véase el interesante artículo DeFeo (2015) sobre las razones que motivan esta situación), como ponen de manifiesto Potowski, Parada y Morgan-Short en su artículo de 2006, quienes apuntaban el dato de que hasta esta fecha solo había un 40% de programas de hablantes de herencia en todo el país (el 67% en California) y de ellos solo un 11% contaba con un examen de clasificación específico para identificar a estudiantes de español como lengua de herencia. Potowski, Parada y Morgan-Short (2006), Fairclough y Ramírez (2009) y DeFeo (2015) son algunos de los autores que llaman la atención sobre la importancia de administrar una prueba de clasificación adecuada para identificar el nivel real de lengua de estos estudiantes y otorgarles una enseñanza específica acorde con sus necesidades educativas para que puedan alcanzar un mayor dominio en los diferentes registros de la lengua.

3.4. Fases de creación de la prueba

En esta sección nos hacemos eco de la propuesta de Bordon (2006: 68) referida a las fases de construcción de un examen de lengua, la cual tuvimos en cuenta para elaborar la prueba de clasificación de NYCCT.

1. Fase previa	2. Fase de diseño
<ul style="list-style-type: none"> • Determinar el objetivo del examen • Precisar el perfil del candidato • Relacionar las tareas del examen en relación con los aspectos lingüísticos contenidos en el Programa de Español • Estimar los recursos necesarios 	<ul style="list-style-type: none"> • Precisar los contenidos: áreas temáticas, nociones y funciones pragmáticas, exponentes lingüísticos (gramática y vocabulario) de acuerdo con un determinado nivel • Establecer los procedimientos: distribución de las partes, tipos de pruebas y duración • Redactar las pruebas y las instrucciones • Definir criterios de evaluación
3. Fase de experimentación	4. Fase de capitulación
<ul style="list-style-type: none"> • Administrar el examen a un grupo de control similar al que está destinado el examen definitivo • Análisis de los resultados 	<ul style="list-style-type: none"> • Comprobar la validez de las pruebas y la fiabilidad de los resultados • Recapacitar para obtener retroalimentación

Figura 2. Propuesta de las fases de construcción de la prueba de clasificación.
Fuente: Bordón, 2006: 68.

En la *fase previa* tuvimos en consideración los aspectos criticados de la prueba administrada con anterioridad, *Avant Assessment* (s.f.), y tratamos de enmendarlos con un nuevo diseño del examen en papel. Para ello, partimos de la necesidad de confeccionar una prueba corta y sencilla que se adaptara al tipo de actividades realizadas en el salón de clase y a los cursos ofrecidos en el Programa de Español en NYCCT. En este sentido, como alternativa a *Avant Assessment*, un primer paso consistió en revisar otro examen comercial estandarizado, la prueba *WebCAPE Spanish Placement Test* (s.f.), creada por la Universidad Brigham. La instrumentalización de esta prueba se basa en el análisis de la gramática y el vocabulario a partir de actividades de selección múltiple con 4 opciones. Sin duda, es un examen fácil de administrar y rápido de resolver para el candidato, que además posibilita el establecimiento de equivalencias de forma sencilla con los temas gramaticales y léxicos incluidos en los cursos de lengua impartidos en NYCCT. A todo ello había que añadir la garantía de que era un procedimiento utilizado por muchas otras de las universidades de CUNY para la clasificación de sus estudiantes de español (y otras lenguas), así como por otras prestigiosas universidades en el Estado de Nueva York y en el resto del país.

Inspirados por la estructura simple y la facilidad de administración de este examen e instigados por la necesidad imperiosa de encontrar una fórmula que se adaptara mejor a la realidad de NYCCT, exploramos diversas opciones y valoramos crear nuestra propia prueba de clasificación en papel siguiendo el modelo de *WebCAPE Spanish Placement Test* (s.f.). Para ello se establecieron primero los objetivos, tipos de actividades y los contenidos lingüísticos que debían incluirse en cada nivel según el currículo vigente en NYCCT.

En la *fase de diseño* se tomó como modelo un examen de diagnóstico utilizado por la Universidad de Salamanca para los ejercicios de gramática y vocabulario. Advertidos de la merma de validez que deriva de rehusar o replicar otros exámenes (Alderson, Clapham y Wall, 1995: 43), decidimos crear nuevos ítems, los cuales

introducían aspectos temáticos contextuales relacionados con la ciudad de Nueva York, y más concretamente con la Universidad de la Ciudad de Nueva York (CUNY). De esta manera, estos ítems se acondicionaron a la realidad educativa de NYCCT con el objetivo de producir una mayor complicidad, receptividad y familiaridad por parte de los estudiantes. A partir de este planteamiento se realizó una primera versión de la plantilla de la prueba de clasificación con 60 preguntas objetivas de gramática y vocabulario. El formato elegido fue el de selección múltiple con aumento progresivo de la dificultad a medida que se avanzaba en el examen. Se establecieron tres bloques de 20 preguntas (30 ítems del nivel inicial, 20 de intermedio y 10 de avanzado alto) que se correspondían con los temas incluidos en los libros de texto utilizados para establecer los niveles de español en NYCCT, *Puntos de Partida* y *En Contacto: Gramática en Acción*. Las instrucciones se tradujeron al inglés para facilitar su procesamiento, debido a que estas no eran el objetivo último de la prueba. En la Figura 3 puede apreciarse la imagen de dos de las preguntas de selección múltiple incluidas en la versión final de la prueba de clasificación:



Figura 3. Preguntas de selección múltiple incluidas en la prueba de clasificación

Fuente: elaboración propia

Con el ejercicio de selección múltiple se suministraba un procedimiento ágil para la resolución de la prueba por parte de los estudiantes y también para su calificación. Junto a la prueba de decisión léxica² propuesta por Fairclough y Ramírez (2009), este es el tipo de tarea objetiva más rápida, fiable y eficaz que conocemos para medir el nivel de lengua de los candidatos. A pesar de que en el ámbito de la enseñanza de lenguas en español han surgido voces muy críticas (Bordón, 2006: 63; Yagüe, 2010: 14-15) con este modelo (argumentando que estas actividades de respuesta cerrada resultan artificiosas y descontextualizadas, debido a que los ítems no proporcionan información pertinente sobre la competencia comunicativa del candidato), otros autores siguen reivindicando el uso de este instrumento por las ventajas que presenta su administración.

Estas virtudes son las que llevan a Brown (1996: 33) a afirmar que cuando un profesor contrario a la filosofía pedagógica subyacente en esta herramienta se convierte en administrador, suele reconocer la utilidad de la prueba para clasificar estudiantes a gran escala y recibir puntuaciones automatizadas a un bajo coste económico. Así, Brown (1996: 33) admite que aunque la evaluación individual de cada candidato es el ideal deseable a alcanzar, se debe asumir que es excesivamente caro en términos tanto de tiempo como de dinero. Sin embargo, también advierte de que los docentes han de ser muy conscientes de los contenidos que se proponen evaluar en términos de currículo

² La prueba de decisión léxica utiliza un instrumento de medición *cloze-test* para medir la habilidad global en un idioma a partir del cálculo del tamaño del vocabulario receptivo que acredita el candidato. Entre sus ventajas, Fairclough y Ramírez (2009) advierten que permite establecer una clara división entre los alumnos que estudian español como una L2 y los que lo hacen como lengua de herencia, aunque reconocen que no resulta efectiva para establecer niveles fiables dentro de este segundo grupo.

y de sus objetivos pedagógicos cuando hacen uso de este instrumento, ya que esta técnica, por ejemplo, no resultaría conveniente para establecer el nivel de dominio de la competencia de destrezas subjetivas, afirmación con la cual estamos de acuerdo. Por el contrario, su empleo para medir el contenido léxico y gramatical de manera fiable ha sido avalado durante varias décadas desde su creación (Bordón, 2015)³.

Sobre el otro tema controvertido –el del enfoque estructural de la tarea–, opinamos que a pesar de los muchos avances que se han sucedido en la metodología de la enseñanza de lenguas en los últimos 50 años (véase Sánchez-Jiménez, 2010), estos elementos lingüísticos siguen siendo el epicentro de la clase de lenguas y de su evaluación. Así lo reconoce también Bordón (2006: 52), quien tras defender la confección de pruebas evaluadoras comunicativas en consonancia con los enfoques de enseñanza más extendidos en la actualidad admite que la gramática y el vocabulario siguen siendo parte central de los exámenes de L2.

y así, aunque hoy día la enseñanza de la lengua se concibe como algo más que la enseñanza de las competencias lingüísticas, sin embargo, sus componentes, gramática y léxico, parecen ser el esqueleto y los músculos que articulan el resto, de manera que siguen constituyendo una parte fundamental de cualquier currículo, programa y manual para el aprendizaje de L2, lo cual tiene su reflejo en la evaluación y en los exámenes. (Bordón, 2006: 259)

Por lo tanto, estos elementos son referentes para la evaluación de la lengua, aunque entendemos –como también señala esta autora– que la gramática y el vocabulario no se pueden medir únicamente de forma aislada (como ocurre con las pruebas objetivas), sino que estos contenidos deben aparecer contextualizados en pruebas de carácter subjetivo, como las de producción escrita e interacción oral integradas en las otras secciones del examen que diseñamos.

Las 8 preguntas cortas creadas para contrastar y constatar los resultados objetivos obtenidos en la primera prueba de selección múltiple también están graduadas según su nivel de dificultad. De esta manera, se caracterizan por el aumento progresivo de su complejidad desde un uso descriptivo de la lengua hasta un nivel más complejo en el que el estudiante debe realizar hipótesis o razonar sobre conceptos abstractos. Esta elevada cantidad de preguntas seleccionadas asegura la recogida de un número considerable de datos disponibles para evaluar el estudiante, a diferencia de lo que ocurría con *Avant Assessment* (s.f.) en el apartado de las pruebas subjetivas (véase § 3.1). Como refleja la Figura 4, las tres preguntas formuladas en el ejemplo –correspondientes al nivel avanzado alto (SPA 2202)– inciden en el uso de formas de futuro, condicional y oraciones compuestas subordinadas, las cuales esperan en este nivel respuestas que contengan estas mismas estructuras o, al menos, argumentos e hipótesis correspondientes a un uso complejo de la lengua expresado mediante otras formas lingüísticas similares.

³ Este procedimiento de análisis objetivo de la lengua fue desarrollado por Kelly en 1915 y se utilizó ampliamente en el modelo psicométrico, el cual se inserta dentro del periodo científico de las etapas de la evaluación descritas por Hinofotis (1981, como se citó en Brown, 1996: 24), correspondientes a las etapas metodológicas de la enseñanza de la lengua (el periodo científico, por ejemplo, establece su correlación con el modelo estructuralista y conductista de la lengua). En su origen, esta prueba nació de la conjunción de los esfuerzos de psicólogos y educadores aunados con el propósito de encontrar una medida objetiva y fiable para valorar los resultados obtenidos sobre la lengua de los candidatos. Con el paso del tiempo, se ha cuestionado su validez debido a que se considera que este tipo de instrumento no es suficiente por sí solo para conocer lo que el estudiante sabe de la lengua o lo que realmente puede hacer con ella en un uso contextualizado, sino que únicamente se centra en medir los elementos constitutivos de la estructura de la lengua (Bordón, 2015: 14).

¿Qué harías tú en esta situación? (What would you do if you were in this situation?) *



Your answer:

¿Volverías a ver esta película? ¿Por qué? (Will you watch this movie again? Why?) *

Your answer:

Si pudieras, ¿qué cambiarías en esta película? (What would you change in this movie if you could?) *

Your answer:

Figura 4. Preguntas cortas de escritura incluidas en la prueba de clasificación
Fuente: elaboración propia

La calificación de esta sección se realiza de forma holística, ya que el grupo de evaluadores del centro es reducido, conocen muy bien los cursos y se coordinan para realizar esta tarea, de modo que pueden adscribir a los estudiantes a un nivel aún prescindiendo de usar una rúbrica analítica explícita. Su cometido es el de asignar directamente un nivel de competencia al candidato a partir de la medición de las 8 preguntas, valorando para ello la capacidad lingüística global del alumno. Esta debe coincidir y corroborar los otros dos métodos empleados en la evaluación, la prueba objetiva de selección múltiple y la de interacción oral.

Este modo de proceder responde a la propia vocación informativa del examen, que sirve al propósito único de designar el curso de español que debe estudiar el candidato en NYCCT. En otras palabras, esta prueba de clasificación local no persigue el objetivo de calificar ni establecer el dominio de lengua del estudiante, función para la cual existen otro tipo de exámenes (Brown, 1996: 3-4; Bordón, 2006: 73-82). A pesar de la falta de precisión en la evaluación del nivel lingüístico de la prueba escrita, esta parte del examen ha resultado especialmente útil en la distinción de estudiantes no nativos de L2 de alumnos para quienes el español es una lengua de herencia, pero fundamentalmente para diferenciar dentro de este segundo grupo a los alumnos que han recibido instrucción formal (literacidad y escritura académica) en la lengua meta (SPA 2202HS) de los que no (SPA 2201HS).

Estas preguntas escritas suponen además un excelente punto de partida para iniciar la conversación en la prueba de verificación oral, en la que se insistirá en los aspectos relacionados con la formación recibida en lengua española y en los contextos

de uso de la lengua a los que el candidato está expuesto a diario (sección I). La segunda parte de esta tarea incluye otro bloque de preguntas (sección II), cuyo objetivo principal reside en valorar el conocimiento que los alumnos poseen de los contenidos lingüísticos evaluados en la prueba. A continuación, se muestra la batería de preguntas propuestas en la sección de interacción oral de la prueba de clasificación:

Sección I. En inglés o en español se le pregunta sobre su bagaje educativo como aprendiente de lenguas:

- ¿Por qué elegiste estudiar español en City Tech?
- ¿Tienes experiencias previas aprendiendo español? Si la respuesta es sí, ¿dónde y cómo lo aprendiste?
- ¿Por cuántos años estudiaste español?
- ¿Cuándo fue la última vez que estudiante español?
- ¿Alguien en tu familia habla español? ¿Quiénes? ¿Hablas con ellos en español?
- ¿Has viajado alguna vez a un país hispanohablante? ¿Cuántas veces? ¿Has vivido allí por algún tiempo?
- ¿Qué otras lenguas hablas? ¿De qué modo las has aprendido?
- ¿Has tomado exámenes de AP en la universidad?

Sección II. En español, con el objetivo de obtener una calificación subjetiva holística de su competencia lingüística

- ¿De dónde eres? ¿Dónde vives?
- ¿Cómo vienes a City Tech? ¿Qué medio de transporte utilizas?
- ¿Qué cursos tomaste el año pasado? ¿Cuántas clases tenías? ¿Tenías español en la escuela secundaria? ¿Hacías algún deporte?
- ¿Qué harás al terminar tus estudios en City Tech? ¿Estudiarás en verano? ¿Viajarás este verano? ¿Vas a trabajar o a estudiar durante el verano? ¿Qué clases tendrás en Fall?
- Si tuvieras la oportunidad, ¿a qué país hispanohablante viajarías? ¿Dónde más te gustaría viajar?
- ¿Qué te gustaría hacer al terminar tus estudios en City Tech? ¿Trabajarías en NYC o fuera de NYC? ¿Por qué?
- ¿Qué quieres hacer en 2020? ¿Qué haras cuando te gradúes? ¿Para qué crees que te puede servir el estudio de una lengua extranjera?
- ¿Qué sugieres que debe hacer una universidad como City Tech para mejorar la calidad de la educación de sus estudiantes?

Figura 5. Batería de preguntas propuestas en la sección de interacción oral de la prueba de clasificación
Fuente: elaboración propia

Evidentemente, tanto el número de preguntas incluidas como la duración final de esta última tarea del examen variará según las dificultades que experimente el evaluador para determinar el curso asignado a cada candidato. Del mismo modo, se podrá modificar esta lista de cuestiones propuesta para ahondar en las partes más relevantes del ejercicio dependiendo del perfil académico y personal del examinando.

La duración total prevista de la prueba en línea (autoevaluación, selección múltiple y expresión escrita) es de unos 30 minutos, lo cual estará condicionado, en parte, por el diverso grado de conocimiento lingüístico que atesoren los candidatos. La entrevista podrá extenderse desde los 5 hasta los 15 minutos. Estas se realizan normalmente en sesiones de una hora cada semana durante la segunda mitad del curso y una semana antes –que se extiende hasta una semana después– de que comience el año académico.

En la *fase de experimentación* los ítems del examen fueron revisados por los 4 profesores que participaron en la administración de la prueba durante su pilotaje. En esta etapa se añadieron comentarios y sugerencias, se corrigieron errores y se mejoraron algunas preguntas de selección múltiple que incorporaban más de una respuesta posible, se modificaron aquellas que podían responderse aplicando el sentido común o las que

contaban con discriminadores engañosos o resultaban ambiguas. Otro factor que tuvimos en cuenta –aprovechando que los 4 profesores pertenecían a diferentes países de habla hispana– fue el del léxico y las construcciones dialectales⁴, que podrían haber dificultado el reconocimiento de la opción correcta por cuestiones no relacionadas directamente con el conocimiento de la lengua. Todos estos elementos fueron identificados y suprimidos en la versión ofrecida a los estudiantes para el pilotaje de la prueba. Del mismo modo, antes de administrar el examen se revisó si el incremento de la dificultad aplicada a la prueba resultaba efectivo y si en cada sección se medían los contenidos impartidos en el programa de los cursos de español ofrecidos en NYCCT.

La prueba se pilotó en NYCCT en su versión en papel durante el tiempo de clase. Un total de 144 estudiantes fueron examinados, divididos en dos cursos por cada nivel en los tres niveles evaluados en la prueba, el inicial (correspondiente al curso SPA 1102), el intermedio (SPA 2201) y el avanzado alto (SPA 3302), de la siguiente manera:

Inicial: *Español 1102* (2 cursos: 28 + 14 =42 estudiantes)
 Intermedio: *Español 2201* (2 cursos: 26 + 20 =46 estudiantes)
 Avanzado alto: *Español 3302* (2 cursos: 28 + 28 =56 estudiantes)

La duración de la tarea varió entre los alumnos, aunque el tiempo total aproximado fluctuó entre los 30-40 minutos. Como era esperable, este periodo resultó ser ligeramente inferior en los grupos de nivel avanzado.

En la *fase de capitulación*, los resultados de la prueba mostraron datos relevantes que permitieron identificar factores inesperados en el análisis, como la falta de compromiso con la prueba (algunos alumnos dejaban sin marcar algunos de los ítems en la actividad de selección múltiple o contestaban a las preguntas de escritura con inesperada brevedad), la gran cantidad de alumnos registrados en un nivel inapropiado y la confirmación del elevado número de hablantes de español como lengua de herencia que concurrían a las clases de español como segunda lengua no solo en el nivel avanzado, sino también en el intermedio. Como resultado de la falta de compromiso con la realización de la prueba mencionada con anterioridad, se prescindió de 6 de los exámenes en la selección final debido a que estos no incluían suficientes datos para proceder al análisis de la prueba de manera fiable).

Los datos se sometieron a revisión estadística para su validación, lo cual confirmó que tanto los ítems establecidos como las tareas diseñadas resultaban adecuados para discriminar los niveles propuestos. Como se aprecia en el Gráfico 2, los estudiantes de nivel avanzado alto (SPA 3302) ofrecieron una respuesta correcta de manera sistemática en las 3 partes en las que se divide el ejercicio de selección múltiple (inicial, intermedio, avanzado alto) en mayor proporción que los del grupo intermedio (SPA 2201), que a su vez obtuvieron un mayor número de acierto con respecto a los alumnos del nivel inicial (SPA 1102), quienes presentaron el índice de acierto más bajo en la prueba.

⁴ En los exámenes de dominio o suficiencia, sin embargo, es recomendable –y hasta esperable– que el contenido dialectal del país donde se administra el examen y en el que se presentan los estudiantes esté representado. Así sucedía claramente en las primeras versiones del DELE, que tenían un impacto más acotado y regional en sus comienzos. Esto cobra mayor sentido en las pruebas locales realizadas en Argentina, los CELU (Certificado de Español Lengua y Uso), y la prueba creada por Ferreira, Vine y Elejalde (2013) en Chile, en los que se incorpora su variedad dialectal.

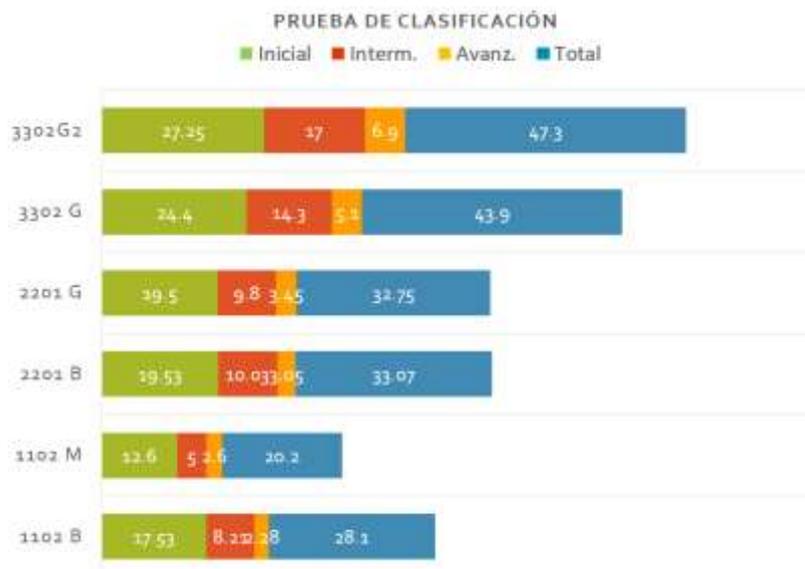


Gráfico 2. Resultados del pilotaje de la prueba de clasificación
Fuente: elaboración propia

Por otra parte, los dos cursos de los niveles intermedio (SPA 2201) y avanzado alto (SPA 3302) recibieron puntuaciones muy similares entre sí, mientras que en el inicial (SPA 1102) se hallaron diferencias significativas, como muestra claramente el gráfico 3. En los gráficos 3 y 4 pueden observarse respectivamente las puntuaciones obtenidas en el total de la prueba y en la primera sección correspondiente al nivel inicial.

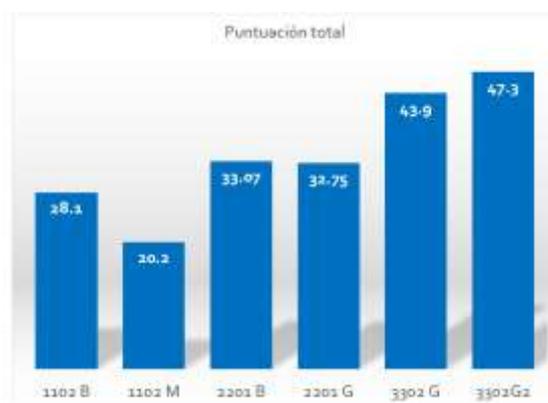


Gráfico 3. Resultados totales obtenidos en el pilotaje de la prueba de clasificación
Fuente: elaboración propia

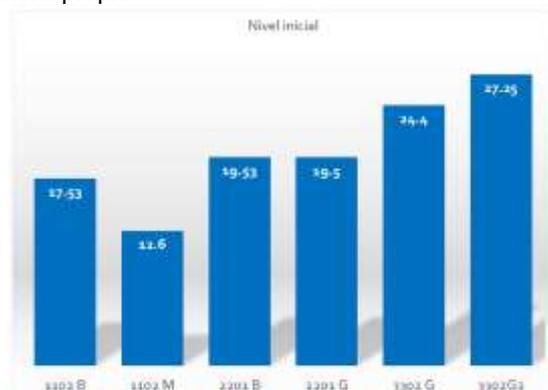


Gráfico 4. Resultados del nivel inicial obtenidos en el pilotaje de la prueba de clasificación
Fuente: elaboración propia

Este mejor conocimiento generalizado de la lengua por parte de uno de los grupos del nivel inicial sobre el otro –reflejado en los Gráficos 3 y 4– puede deberse principalmente a dos factores: que el conjunto de estudiantes en una de las clases poseía un mejor conocimiento de la lengua; también podía deberse a que la instrucción había sido menos efectiva en una de las clases y, en consecuencia, los alumnos mostraron de manera generalizada un aprendizaje más deficiente de los contenidos lingüísticos estudiados durante el semestre (teniendo en cuenta que el pilotaje se realizó justo antes de la finalización del curso).

Las respuestas a las preguntas escritas corroboraron los datos obtenidos en la prueba objetiva. También sirvieron para identificar la amalgama de niveles dentro de la misma clase, como en los dos ejemplos que se muestran a continuación, en los que los estudiantes de español como lengua de herencia habían sido acomodados en cursos inferiores a su correspondiente nivel de lengua. Este primer ejemplo pertenece a un estudiante matriculado en el nivel inicial (SPA 1102), equivalente a un nivel A2 de acuerdo con el Marco Común Europeo de Referencia para las Lenguas (MCER).

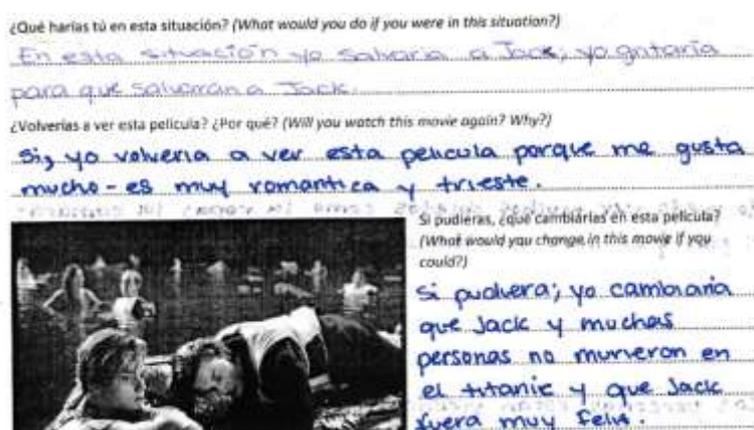


Figura 6. Respuestas de las preguntas escritas de un estudiante del curso SPA 1102 en el pilotaje de la prueba de clasificación

Fuente: elaboración propia

El segundo ejemplo pertenece a un alumno del nivel intermedio (SPA 2201), correspondiente al nivel B1-B1+ del MCER.

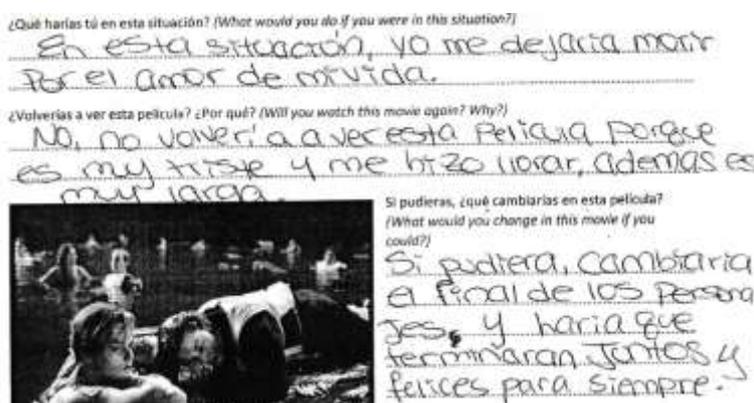


Figura 7. Respuestas de las preguntas escritas de un estudiante del curso SPA 2201 en el pilotaje de la prueba de clasificación

Fuente: elaboración propia

La fase de pilotaje sirvió también para eliminar aquellos ítems de la prueba objetiva que no habían discriminado entre los niveles a causa del bajo índice de aciertos identificados en los niveles inicial e intermedio o, al contrario, por resultar excesivamente fáciles para ser incluidos en la sección del nivel avanzado alto. Los nuevos ítems creados para reemplazarlos fueron incluidos en la prueba revisada y serán validados nuevamente mediante las respuestas obtenidas en el examen en línea cuando se hayan analizado los datos de los dos próximos semestres, ya que *Google Forms* permite computar el número de aciertos del total de los candidatos por ítem. Será entonces cuando se valore la fiabilidad de los resultados obtenidos en el pilotaje, lo que servirá una vez más para rediseñar las tareas en caso de que esto fuera preciso.

Después de los dos próximos cursos académicos (Primavera 2020 y Otoño 2021) nos planteamos validar igualmente la prueba de clasificación en línea con el fin de identificar posibles errores y determinar los siguientes pasos a seguir en su remodelación y en el ajuste de las tareas. Un estudio que inspira esta decisión es el de Johnson y Riazi (2017: 6). Estos autores comprobaron en encuestas realizadas tras la creación y administración de su examen que el 37% de los estudiantes no superó el curso en el que habían sido ubicados y el 15% de ellos lo encontró demasiado difícil, lo que probaba que el examen desarrollado no había resultado efectivo. Los datos de este estudio también revelaron que el 25% había experimentado serios problemas para completar la prueba y que el 22% consideraba que el tiempo asignado resultaba insuficiente. Nos parece interesante realizar también este tipo de investigación en NYCCT, ya que somos conscientes de que la administración de un examen de clasificación inadecuado puede tener consecuencias negativas tanto para el alumnado (como suspender el curso o dedicar un extenuante sobreesfuerzo para superarlo) como para la institución (relacionados con la retención y el progreso de los estudiantes en sus respectivas carreras).

Hasta el momento, la nueva prueba de clasificación de español diseñada para NYCCT ha servido para ubicar a más de 500 estudiantes durante el semestre de la primavera de 2019 y otoño de 2020. Los resultados obtenidos en esta primera aplicación del examen invitan al optimismo, pues sólo se han producido 6 cambios de nivel de los estudiantes matriculados en los cursos de español, lo que acredita la validez predictiva de la prueba. A pesar de ello, estos resultados preliminares no han sido confirmados aún con datos cualitativos sobre la percepción de los propios estudiantes sobre la prueba, más allá de los comentarios realizados en conversaciones informales previas a la entrevista de verificación del examen en línea. Estas explicaciones permitieron identificar algunas de las ventajas de la prueba, como son –insistimos, en opinión de los candidatos– la facilidad de navegación, el tiempo empleado en su ejecución, la especificidad de las preguntas planteadas y la progresión en el nivel de dificultad.

4. Conclusiones

La nueva prueba de clasificación para agrupar a los estudiantes de español en NYCCT fue creada como consecuencia de las necesidades específicas de nuestra institución y las particulares características que presentan nuestros alumnos. El objetivo pedagógico de esta herramienta tenía como fin establecer una medida fiable de los contenidos lingüísticos aprendidos por los estudiantes en sus experiencias educativas previas o a través de su experiencia vital en contextos naturales de adquisición de la lengua. El diseño de este examen ha permitido superar los problemas identificados en las pruebas de clasificación que fueron utilizadas previamente en NYCCT, consiguiendo como

resultado un mayor equilibrio en el nivel de los estudiantes asignados en cada curso. Entre estas mejoras se encuentran la reducción del tiempo de duración y la flexibilidad en la administración del examen, el aumento de la fiabilidad motivado por el uso de diferentes metodologías complementarias (objetivas y subjetivas) para establecer el análisis de la prueba y la atención personalizada prestada a la situación lingüística y académica de los candidatos en la parte de autoevaluación y en la entrevista oral de verificación del nivel asignado.

El pilotaje de la prueba sirvió para identificar a un número significativo de estudiantes que no estaban propiamente ubicados en su nivel, además de esclarecer algunos errores en la fiabilidad de los ítems que fueron inmediatamente corregidos. El primero de estos hallazgos permitió realizar cambios en el currículo del centro mediante la creación de un nuevo curso de lengua (SPA 2202HS), lo cual contribuyó a la gestación de una nueva secuencia en el Programa de Español para responder a las necesidades pedagógicas específicas que presentan los hablantes de español como lengua de herencia.

En base a los datos obtenidos en la fase de validación, advertimos que el examen también puede resultar útil para diagnosticar a aquellos estudiantes que están experimentando problemas con el aprendizaje de los contenidos del curso antes de que este finalice. Como consecuencia, esta herramienta podría utilizarse a lo largo del semestre como medida para predecir los casos de fracaso académico y posibilitar al alumnado la mejora de su actualización en el curso antes de la finalización del semestre.

Con respecto a las carencias, una de las mayores limitaciones observadas durante el pilotaje ha sido la relacionada con la parte técnica de la prueba. La plataforma de *Google Forms*, aunque se mostró solvente en el propósito de lograr los objetivos planteados con el diseño del examen, resultaba excesivamente limitada en las prestaciones de algunos aspectos del *software*. De esta manera, por ejemplo, hubiera sido deseable haber podido realizar las tareas del examen utilizando herramientas visuales y sonoras que incorporaran un mayor grado de animación en el ejercicio y que sirvieran para contextualizar más apropiadamente los contenidos lingüísticos expuestos en la prueba asociados a un tipo de situación de uso real de la lengua. Por este motivo, en el futuro pensamos explorar otros tipos de *software* más sofisticados para alojar la prueba de clasificación, como el creado por otra escuela de CUNY, Lehman College (Durán-Urrea y Lado, 2015), en conjunción con una empresa de tecnologías de la información. Su examen permite medir las producciones de los estudiantes de herencia de español con el objetivo de discriminar entre los distintos niveles que se ofrecen en esta institución.

Esperamos que el modelo mostrado en este artículo sobre la creación de una prueba de clasificación local inspire a otros centros a seguir este ejemplo, pues como se ha expuesto a lo largo del texto, este examen presenta una serie de ventajas de adaptabilidad en la medición de los contenidos lingüísticos frente a las pruebas estandarizadas de lengua, las cuales, sin duda, permiten identificar inconsistencias en el diseño y mejoran tanto el rendimiento del examen como los resultados de clasificación de los grupos de manera específica en cada institución.

Referencias bibliográficas

- ALDERSON, Ch. J., CLAPHAM, C. y WALL, D., 1995. *Language test construction and evaluation*. Cambridge, Cambridge University Press. ISBN 9780521478298.
- AIRE. Data Dashboard. Enrollment Trends. <<http://air.citytech.cuny.edu/data-dashboard/>> [2 nov. 2019]
- AVANT ASSESSMENT SPANISH, s.f. Disponible en: <<https://avantassessment.com>> [12 sept. 2019]
- BACHMAN, L. F. y PALMER, A.S., 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford, Oxford University Press. ISBN 9780194371483.
- BARALO, M., 2012. “Diseño de certificación lingüística comunicativa: ¿cómo evaluar la expresión e interacción escritas y orales?”, *Revista Internacional de Lenguas Extranjeras*, 1, pp. 11-30.
- BEAUDRIE, S., DUCAR, C. y POTOWSKI, K., 2014. *Heritage language teaching: Research and practice*. New York, McGraw Hill. ISBN 9781308299044.
- BORDÓN, T., 1990. Evaluación de la lengua hablada. Antología de los encuentros internacionales del español como lengua extranjera. *marcoELE*, 11. ISSN 1885-2211. Disponible en: <<http://marcoele.com/descargas/navas/16.bordon.pdf>> [12 feb. 2019]
- BORDÓN, T., 2006. *La evaluación de la lengua en el marco del E/L2: bases y procedimientos*. Madrid, ArcoLibros. ISBN 8476356447.
- BORDÓN, T., 2015. “La evaluación de segundas lenguas (L2). Balance y perspectivas”, *Revista Internacional de Lenguas Extranjeras*, 4, pp. 9–30.
- BROWN, J. D., 1996. *Testing in Language Programs*. New Jersey, Prentice Hall. ISBN 9780131241572.
- CIARRA TEJADA, A., 2013. “La prueba inicial para el diagnóstico previo de nivel y la organización de grupos. Modelo y descripción para ELE”, *Revista redELE*, 25, pp. 1-20.
- CONSEJO DE EUROPA, 2001. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: MECD-Anaya. <http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf> [15 sept. 2018]
- DEFEO, D. J., 2015. “Spanish is foreign: Heritage speakers’ interpretations of the introductory Spanish language curriculum”, *International Multilingual Research Journal*, 9(2), pp. 108-124.
- DORWICK, T., PÉREZ-GIRONÉS, A.M. y BECHER, A., 2016. *Puntos de partida*. (10a ed.). New York, McGraw Hill. ISBN 9780073534497.
- DURÁN-URREA, E., y LADO, B., 2015. Placement of Heritage Language Learners of Spanish at the College Level. En: *25th Conference on Spanish in the United States and 10th Spanish in Contact with Other Languages* [Comunicación]. New York: The City College of New York, CUNY, 28 de marzo, 2015.
- EDELE, A., SEURING, J., KRISTEN, C. y STANAT, P., 2015. “Why bother with testing? The validity of immigrants’ self-assessed language proficiency”, *Social Science Research*, 52, pp. 99-123.

- FAIRCLOUGH, M., y RAMIREZ, C. J., 2009. “La prueba de decisión léxica como herramienta para ubicar al estudiante de español en los programas universitarios”, *Íkala, revista de lenguaje y cultura*, 14(21), pp. 85-99.
- FERNÁNDEZ, S., y STAIRE, S., 2017. “Estrategias docentes en el aula de ELE. Y tú ¿cómo lo haces, cómo evalúas, cómo trabajas con tareas, cómo...?”, En AINCIBURU, M. C. y FERNÁNDEZ SILVA, C., *La adquisición de la lengua española: aprendizaje, enseñanza, evaluación*, Buenos Aires, Editorial Autores de Argentina, s.p.
- FERREIRA, A., VINE, A. y ELEJALDE, J., 2013. “Hacia una prueba de nivel en Español como Lengua Extranjera”, *Revista de Lingüística Teórica y Aplicada*, 51(2), pp. 73-103.
- FERRIS, D., EVANS, K. y KURZER, K., 2017. “Placement of multilingual writers: Is there a role for student voices?”, *Assessing Writing*, 32, pp. 1-11.
- GARCÍA LABORDA, J. y NAVARRO, C., 2008. “Automatizar los exámenes del Diploma de Español como Lengua Extranjera (DELE): ¿Utopía o realidad?”, *Revista de Lingüística Teórica y Aplicada*, 46(2), pp. 81-93.
- GILL, M., WEGMAN, B. y MENDEZ-FAITH, T., 2015. *En contacto: Gramática en acción*. (9ª ed.). Boston, Cengage. ISBN 9780495912651.
- JOHNSON, R. y RIAZI, A.M. (2017). “Validation of a locally created and rated writing test used for placement in a higher education EFL program”, *Assessing Writing*, 32, pp. 84-105.
- KIM, H. R., BOWLES, M. A., YAN, X. y CHUNG, S. J., 2018. “Examining the comparability between paper-pencil and computer-mediated versions of an integrated writing placement”, *Assessing Writing*, 36, pp. 49-62.
- LONG, A., SHIN, S., GEESLIN, K. y WILLIS, E., 2018. “Does the test work? Evaluating a web-based language placement test”, *Language Learning & Technology*, 22(1), pp. 84-94.
- MARTÍN LERALTA, S., 2009. “Certificación lingüística de nivel inicial para inmigrantes en contexto laboral: ejemplo de una prueba del examen DILE”, *Lengua y migración*, 3(1), pp. 89-104.
- ORTEGA, L., 2019. “SLA and the Study of Equitable Multilingualism”, *The Modern Language Journal*, 103, pp. 23-38
- OTHEGUY, R. y ZENTELLA, A., 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford, Oxford University Press. ISBN 9780199737406.
- PEREZ-GUILLOT, C, ZABALA, J. y JAIME, A., 2013. “Computer-based Language Tests in a University Language Centre”, *Language testing and Evaluation*, 28, pp. 145-161.
- POTOWSKI, K., PARADA, M. y MORGAN-SHORT, K., 2012. “Developing an online placement exam for Spanish heritage speakers and L2 students”, *Heritage Language Journal*, 9(1), pp. 51-76.
- SÁNCHEZ-JIMÉNEZ, D., 2010. La enseñanza de lenguas extranjeras en el entorno europeo. *Enciclopedia Virtual de Elenet*. ISSN 1772-9955. Disponible en: <<http://www.elenet.org/Publicaciones/Enciclopedia/La-enseñanza-de-lenguasextranjeras-en-el-entorno-europeo/bibliografia2.asp?palabra=2>> [10 nov. 2014]

THE CITY UNIVERSITY OF NEW YORK (CUNY). Academics. Disponible en: <<https://www.cuny.edu/academics/>> [2 nov. 2019]

VALDÉS, G., 1997. “The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions”, en COLOMBI, M. C. y ALARCÓN, F. X. (Eds.), *La enseñanza del español a hispanohablantes*. Boston, MA: Houghton Mifflin, pp. 8-44.

webCAPE (Computer Adaptive Placement Exam), Spanish., s.f., Orem, UT: Perpetual Technology Group. Disponible en: <<http://www.perpetualworks.com>> [17 oct. 2019]

YAGÜE, A. (2010). Examinando exámenes. Consideraciones acerca de las pruebas de medición en ELE. *marcoELE*, nº 10. ISSN 1885-2211. Disponible en: <http://marcoele.com/descargas/10/yague_examinandoexámenes.pdf> [12 sept. 2018]