

Creación, Desarrollo y Resultados de la Aplicación de Pruebas de Evaluación basadas en Estándares para Diagnosticar Competencias en Matemática y Lectura al Ingreso a la Universidad

Creation, Development and Assessment Tests Standards-based Application Results to Diagnose Math and Reading Skills to University Entrance

Pilar Rodríguez Morales *

Universidad de la República

El artículo presenta el proceso de creación y desarrollo de pruebas de evaluación de las competencias en Matemática y Lectura para el nivel de ingreso a la Universidad y su aplicación a los ingresantes a los Centros Universitarios Regionales de la Universidad de la República (Uruguay) que atraen a un perfil de estudiantes con vulnerabilidad social y académica. La metodología utilizada para el desarrollo de las pruebas implicó la creación y establecimiento de estándares de contenido a través de grupos de expertos. La clasificación en los estándares de desempeño se obtuvo a través de un método nuevo propuesto por García et al. (2013) basado en la Teoría de Respuesta al Ítem (TRI). Se obtuvieron dos pruebas unidimensionales, con muy buena consistencia interna, validadas y calibradas mediante TRI que clasificaron a los estudiantes en tres niveles de desempeño (insuficiente, suficiente y avanzado). El 22% superó la suficiencia en la prueba de Matemática y el 53% lo logró en la prueba de Lectura. Aunque la proporción de estudiantes que alcanzaron la suficiencia es baja, estos resultados son coherentes con otras investigaciones e impulsaron líneas de acción que incluyeron la creación de grupos académicos en Lectura y Matemática para el diseño de programas de apoyo.

Palabras clave: Pruebas diagnósticas, Prueba de matemática, Prueba de lectura, Ingreso a la universidad, Estándares académicos.

The article presents the process of creation and development of assessment tests in Math and Reading skills for entry level to university and its application to entrants to Regional University Centers of the University of the Republic (Uruguay) that attract a profile of students with social and academic vulnerability. Methodology used for tests development involved creation and establishment of content standards through expert groups. Performance standards' classification was obtained through a new method proposed by Garcia et al. (2013) based on the Item Response Theory (IRT). Two one-dimensional tests were obtained, with very good internal consistency, validated and calibrated by IRT that classified students in three performance levels (insufficient, sufficient and advanced). The 22% overcame the sufficiency in Math test and 53% it reached in Reading test. Although students' proportion of achieving proficiency is low, these results are consistent with other research and it promoted action lines that included the creation of academic groups in Reading and Mathematics for the design of support programs.

Keywords: Diagnostic tests, Mathematics tests, Reading tests, College admission, Academic standards.

*Contacto: prodriguez@cure.edu.uy

issn: 1989-0397

www.rinace.net/riee/

<https://revistas.uam.es/riee>

Recibido: 26 de octubre de 2016

1ª Evaluación: 2 de diciembre de 2016

Aceptado: 22 de diciembre de 2016

1. Introducción¹

El ingreso a la Educación Superior supone una serie de retos que los estudiantes deben enfrentar, siendo la exigencia académica uno de los que plantea mayores dificultades. Por este motivo, el desempeño de los estudiantes de ingreso a la universidad se ha convertido en objeto de diversas investigaciones (Attewell, Heil y Reisel, 2012).

Los sistemas universitarios latinoamericanos tienen procesos de ingreso muy disímiles entre sí, encontrándose algunos sistemas con una selección muy competitiva y estricta hasta otros con acceso libre e irrestricto como es el caso del Uruguay. La Universidad de la República de Uruguay (Udelar), la principal institución universitaria del país, habilita el ingreso a todos los que acrediten haber finalizado el nivel educativo previo. Por otro lado, la educación media superior (bachillerato) tampoco realiza pruebas de evaluación sistemáticas sobre los aprendizajes de sus egresados, la última se realizó en 2003. Se torna necesario, entonces, la evaluación diagnóstica de las competencias al ingreso a la Universidad, ya que acceden a la educación superior estudiantes con diferentes niveles académicos. La información proporcionada por este tipo de evaluaciones es de gran utilidad para la planificación de acciones de mejora y la orientación de las trayectorias académicas de los estudiantes.

La Educación Superior de Uruguay no cuenta con una extensa tradición en realizar pruebas estandarizadas, como existe en otros niveles del sistema educativo, principalmente Primaria. Sin embargo, la Udelar ha desarrollado pruebas de evaluación de tipo diagnóstico para los estudiantes de ingreso desde las distintas facultades (Altmark et al., 2006; Enrich et al., 2006; Míguez, Blasina y Loureiro, 2013; Mussio y Martinotti, 2013; UEFI, 2012; Unidad de Enseñanza de Facultad de Ciencias, 2005, 2010). El principal objetivo de estas evaluaciones es conocer las competencias de los estudiantes que ingresan en las áreas consideradas básicas para la carrera en que se matricularon (matemática, física, química y comprensión lectora) con el fin de facilitar la transición entre ciclos educativos. No obstante, todas estas experiencias se han circunscripto a una facultad o área, no se basaron en estándares preestablecidos, el análisis de los datos ha incluido solamente teoría clásica de los tests y no se ha utilizado un método para el establecimiento del punto de corte.

Por otra parte, el proceso de descentralización llevado adelante por la Udelar desde 2007 que tuvo como corolario la creación de siete centros universitarios en el interior del país, es decir fuera de Montevideo, donde históricamente se concentraron las actividades universitarias, crea oportunidades de acceso a un perfil de estudiantes con ciertas vulnerabilidades, tanto sociales como académicas. Estos estudiantes reúnen características asociadas a la no culminación de los estudios universitarios, tales como dificultades para la finalización educación media, los antecedentes académicos del núcleo familiar y las competencias adquiridas (Rodríguez, 2014). En este sentido, se torna vital conocer las competencias de los estudiantes al ingreso para enfrentar un nivel educativo

¹ Agradecimientos: A los Dres. Ma. Ángeles González Galán y Tabaré Fernández Aguerre directores de la tesis doctoral de la autora "Creación y establecimiento de estándares para la evaluación de la calidad de la educación superior: un modelo adaptado a los Centros Universitarios de la Udelar", defendida en la UNED, donde se produjeron parte de los resultados que se incluyen en este artículo y a los revisores de este artículo por sus valiosos comentarios y sugerencias para mejorarlo.

superior con el fin de adaptar los programas educativos y crear dispositivos de apoyos específicos. Frente a la necesidad de evaluar las competencias de los estudiantes que ingresan a estos centros, se desarrollaron diversas pruebas con este fin (Rodríguez y Correa, 2011; Rodríguez, Correa y Díaz, 2012; Rodríguez, Díaz y Correa, 2013, 2014).

Todas las pruebas referenciadas hasta aquí se basaron en la evaluación de ciertos contenidos, que en algunos casos fueron acordados por los profesores. Sin embargo, se consideró necesario crear pruebas basadas en estándares donde se evaluarán los aprendizajes logrados por los estudiantes, se establecieran metas educativas, y por tanto, que su evaluación fuera coherente con los estándares fijados (Linn y Gronlund, 2000; O'Shea, 2005; Hamilton, Stecher y Yuan, 2008). Ferrer (2006) enfatiza la necesidad de establecer estándares aun cuando existan marcos curriculares nacionales como forma de regular la forma en que se realiza la cobertura de los contenidos y, además, que los criterios de evaluación no resulten dispares entre diferentes contextos.

Por las razones expuestas anteriormente, desde 2014, se comienza a desarrollar pruebas de evaluación basada en estándares, que se aplican por primera vez a la cohorte de estudiantes que ingresaron al siguiente año (Rodríguez et al., 2015). La competencia matemática ha sido considerada como indispensable para los estudiantes que ingresan a la mayoría de las titulaciones y la competencia en lectura se considera transversal a todos los programas educativos. Además, matemática y lectura son consideradas competencias básicas para un estudiante de ingreso a la Universidad (Bertoni, 2005; Zalba et al., 2005). Por estos motivos, se propuso desarrollar pruebas diagnósticas basadas en estándares en estas dos áreas.

El objetivo principal de este artículo es mostrar el proceso de creación y desarrollo de pruebas para la evaluación de competencias básicas (matemática y lectura) al ingreso a la Universidad basadas en estándares, previamente acordados por grupos de expertos, donde se utilizó un método basado en la Teoría de Respuesta al Ítem para el establecimiento de los estándares de desempeño. Además, se presentarán los principales resultados obtenidos a través de estas pruebas y las acciones implementadas para mejorar las carencias detectadas.

2. Fundamentación teórica

El objetivo de las pruebas que nos propusimos elaborar es determinar la posición de los estudiantes en relación con el constructo definido y no su clasificación en función de los individuos que conforman la cohorte. Por este motivo, las pruebas desarrolladas son referidas al criterio y no a la norma. Estas últimas no proporcionan la información necesaria para el diagnóstico de habilidades cognitivas, mientras que el enfoque utilizado se fundamenta en que, a través de los métodos seleccionados, se determine con precisión las competencias de los estudiantes de ingreso. Por eso, el modelo de evaluación elaborado tiene en cuenta una determinada estructura cognitiva, asociada a sus marcos teóricos respectivos –para Matemática el desarrollado por The College Board (2014) y para Lectura el desarrollado por ANEP (2011)- y un conjunto de criterios psicométricos para la elaboración de las pruebas.

La evaluación de tipo diagnóstico en el ámbito educativo tiene su origen en la necesidad de caracterizar a los estudiantes y obtener información que puede ser interpretada por los tomadores de decisiones (Rupp, Templin y Henson, 2010). El principal objetivo es

medir el atributo que se pretende evaluar, por eso se busca desarrollar evaluaciones que sean capaces de determinar con precisión las habilidades cognitivas a evaluar (Gitomer, Steinber y Mislevy, 2009). En nuestro caso, el objetivo de las pruebas diagnósticas desarrolladas es determinar en qué grado los estudiantes superan el constructo definido como competencia en el área de matemática y lectura.

La creación de estas pruebas está sostenida en tres conceptos fundamentales: las pruebas criterioles, los estándares y la interpretación de las puntuaciones basada en la Teoría de Respuesta al Ítem (TRI).

2.1. Pruebas criterioles

Las pruebas referidas al criterio son construidas para medir un determinado dominio de aprendizaje y de esta forma situar a los individuos en relación con ese dominio (Pérez Juste, 2006). Se las diferencia de las pruebas normativas porque mientras estas tratan de ubicar la posición relativa del sujeto con respecto a los demás, las pruebas criterioles tratan de ver en qué medida domina el criterio de referencia (Muñiz, 1998) y permiten conocer con qué grado un estudiante alcanza los niveles de aprendizaje preestablecidos.

El origen de las pruebas criterioles se encuentra en el trabajo de Glaser (1963) donde las define como tests elaborados para establecer el nivel de ejecución de un examinado con respecto a un dominio bien definido. Por eso, para la construcción de una prueba criterial debe estar bien definido el universo de medida. El dominio constituye el conjunto de indicadores apropiados para representar el nivel de los sujetos en el constructo que se concretan en ítems (Prieto y Delgado, 1996). Para la creación o adaptación de los ítems se recomienda realizar una revisión desde una perspectiva lógica –a través de jueces que analicen la coherencia entre el ítem y el objetivo a medir– o empírica –la calidad técnica del ítem– (Pérez Juste, 2006).

El desarrollo de pruebas criterioles se fundamentan en dos temas: el análisis y la especificación del dominio y el desarrollo de estándares o un sistema de interpretación de las puntuaciones (Jornet y González Such, 2009).

Según Gil Pascual (2016) las pruebas criterioles se clasifican en función del dominio instruccional de referencia en: test de certificación o admisión (tienen como objetivo decidir sobre la consecución del dominio instruccional de un nivel educativo), test de nivel o dominio (se utilizan para decidir sobre el nivel de instrucción para promover de un curso a otro), test de aula o clase (se circunscribe al aula), test de diagnóstico (pretenden detectar la existencia de problemas relacionados con el aprendizaje), test individualizado (recoge información de aspectos procesuales o de rendimiento de un sujeto).

Las pruebas diagnósticas que se desarrollaron tienen como objetivo determinar con la mayor exactitud posible cuáles son las competencias cognitivas de los estudiantes, por eso se utilizan métodos que pueden aportar esa precisión. Los resultados de estas pruebas se utilizaron para adaptar la enseñanza y el aprendizaje a las necesidades de los estudiantes (Black y Wiliam, 1998) y proveer información para los tomadores de decisiones con el objetivo de planificar políticas educativas.

2.2. Estándares

En las últimas décadas del siglo XX se generaliza la utilización de estándares asociados a la evaluación de aprendizajes (Glass, 1978), especialmente su uso en las pruebas

criteriales (Burton, 1978). En América Latina se comienza a estudiar la relación entre los currículos y las pruebas criteriales a fines de los 90, cuando se conforman grupos de trabajo sobre la temática que recomiendan que los gobiernos establezcan estándares educativos y desarrollen pruebas para medir los resultados (Ferrer, Valverde y Esquivel, 2006).

El concepto de estándares, entendidos en un sentido amplio, se define como un criterio fijo respecto del cual se juzga el resultado o también puede ser comprendido como el logro obtenido (Sotomayor y Gysling, 2011).

En el marco de este trabajo, entendemos los estándares como descriptores de logro, diferenciados por niveles, técnicamente definidos y previamente acordados. Un descriptor debe operacionalizarse para ser evaluado, puede hacerse en términos dicotómicos (presencia/ausencia o disponibilidad/carencia), ordinales (gradación, por ejemplo de acuerdos) o en puntajes (variables métricas).

Un aspecto importante del establecimiento de estándares es alcanzar acuerdos entre los implicados para que sean validados por la comunidad en la que van a ser aplicados. Cizek y Bunch (2007) diferencian entre estándares de contenido y estándares de desempeño. Mientras los primeros son un conjunto de resultados, objetivos curriculares o metas específicas de instrucción que forman el dominio desde el cual se construye un test, el segundo concepto es usado como sinónimo de punto de corte, nivel de logro.

Los estándares de contenido son definidos como la descripción de conocimientos o habilidades específicas sobre la que se espera que los examinados demuestren su dominio acorde a su edad, nivel o campo de estudio (Cizek, Bunch y Koons, 2004) o lo que los estudiantes deben saber y saber hacer en determinadas áreas (Tourón, 2009), es decir los logros de aprendizaje, para demostrar o acreditar ciertos niveles de aprendizajes o competencia. También estos descriptores son tomados como parte de la validez de contenido del instrumento (Hambleton, 2001). Esta diferenciación entre estándares de contenido y estándares de desempeño o rendimiento ha sido la más ampliamente difundida y adoptada (Ravitch, 1996; Tourón, 2009).

Los estándares de desempeño se definen como la descripción del grado de desempeño de los examinados en diferentes categorías (Cizek, Bunch y Koons, 2004) y son usados la mayoría de las veces para informar sobre el desempeño de grupos de estudiantes y el progreso de los centros educativos o estados, en vez de ser utilizados para tomar decisiones acerca de los estudiantes individuales (Linn, 2003).

El objetivo del establecimiento de estándares de desempeño es la clasificación de los estudiantes en niveles. La descripción de los niveles de desempeño provee información sobre los niveles establecidos. Esta descripción es un listado de conocimientos, habilidades o atributos que se consideran integran el nivel de desempeño y que pueden variar en su especificidad (Cizek y Bunch, 2007). Pueden ser planteados con anterioridad al establecimiento de estándares de desempeño (punto de corte) o pueden ser elaborados por el mismo grupo que establezca los estándares de desempeño (Lewis y Green, 1997; Mills y Jaeger, 1988). Ejemplos de descripciones de estándares de desempeño pueden consultarse en Cizek y Bunch (2007), Jornet y González Such (2009) o Tourón (2009).

2.3. Teoría de respuesta al ítem

Con el fin de optimizar la medición de constructos psicológicos o cognitivos y por lo tanto, mejorar la toma de decisiones y resolver ciertos problemas de medición, se ha pasado de la utilización de la Teoría Clásica de los Tests (TCT) a la Teoría de Respuesta al Ítem (TRI). La TRI toma los ítems como unidad de análisis, permite describir algunas propiedades psicométricas del instrumento mediante indicadores invariantes, es decir, que no dependen de la muestra en que se aplique. Esa se puede considerar como su mayor contribución (Muñiz, 1997).

La TRI propone soluciones para las limitaciones de la TCT como la invarianza de los parámetros que permite que el valor de los parámetros de los ítems no depende de la muestra de donde se obtiene; la precisión con la que cada persona es medida según su nivel de rasgo y en función de los ítems concretos que se le hayan aplicado; y los indicadores de bondad de ajuste que permiten estudiar el grado en que los datos ajustan al modelo (Abad et al., 2011).

La TRI se diferencia de la teoría clásica de los tests por utilizar modelos basados en las características de los ítems en vez de las del test, donde las características de los ítems son independientes del grupo en el ítem se ha calibrado y las puntuaciones del rasgo no dependen de las puntuaciones obtenidas en cada test particular. En la TRI se puede obtener una medida de la precisión para cada puntuación del rasgo, lo que la distingue claramente de la teoría clásica y para evaluar la fiabilidad no se requieren de tests estrictamente paralelos.

La TRI establece una relación funcional entre la respuesta del examinado a cada ítem y el rasgo latente responsable de tal realización, es decir, la habilidad, que se nota θ . En la mayoría de los modelos se asume que esta función depende solo de un rasgo, esto es, son unidimensionales. La función que da la probabilidad de obtener determinada puntuación en el ítem condicionado al rasgo, se denomina curva característica del ítem. Las hipótesis que sustentan la TRI son: la unidimensionalidad del espacio latente, la independencia local y la ausencia de factores de velocidad.

3. Métodos

El proceso de trabajo que dio lugar al establecimiento de estándares y creación de instrumentos para evaluar las competencias en Matemática y Lectura se desarrolló durante tres años y se muestra en las figuras 1 y 2.

Se partió de la revisión de los programas curriculares de Matemática de los últimos dos años de bachillerato, estableciéndose como referencia de competencia mínima los contenidos de los programas de las orientaciones de bachillerato con menos contenidos de Matemática. Un grupo de docentes elaboró estándares de contenido agrupados en cuatro categorías: números y operaciones, álgebra y funciones, geometría y análisis de datos, estadística y probabilidad. Esta selección de estándares se presentó a un grupo de discusión integrado por docentes de Matemática de nivel universitario y de Educación media, para ser evaluada. Se realizaron varias rondas de consultas hasta que se acordaron los estándares de contenido sobre los que se basaría la prueba.

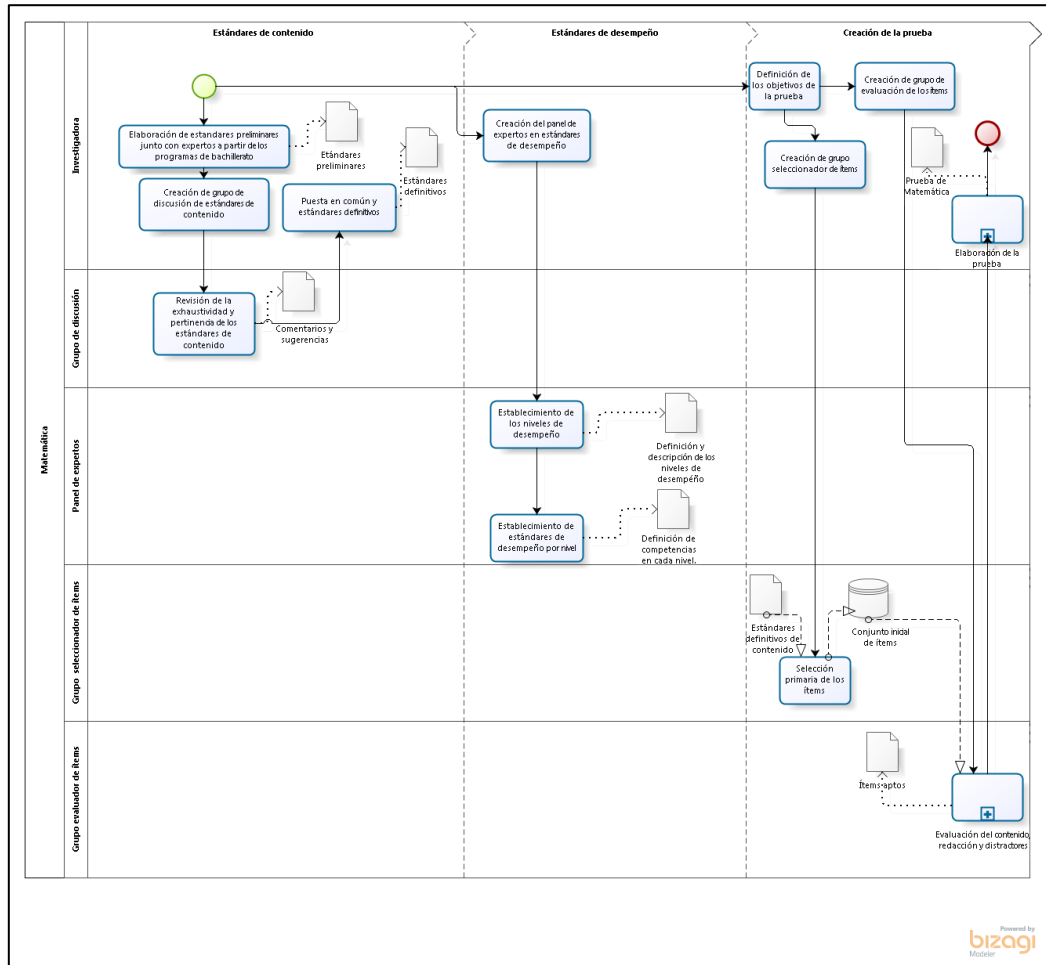


Figura 1. Proceso de establecimiento de estándares y creación de las pruebas de matemática

Fuente: Elaboración propia.

Un panel de expertos, conformado por algunos de los participantes del grupo de discusión que estableció los estándares de contenido más otros especialistas en evaluación de aprendizajes en Matemática, acordó establecer tres niveles de desempeño, que se denominaron Insuficiente, Suficiente y Avanzado y establecieron las competencias que debe tener un estudiante para alcanzar cada uno de los niveles (Rodríguez, 2016). Con la finalidad de contar con ítems ya testeados, se eligió tomar como base la dimensión Matemática del *Scholastic Assessment Test* (SAT) elaborada por The College Board (2014). Un grupo de docentes preseleccionó 186 ítems, repartidos equilibradamente entre las diferentes categorías. Con el objetivo de evaluar la adecuación de los ítems a los estándares de contenido se conformó un grupo de discusión integrado por docentes con experiencia en Educación Media Superior y Universidad y en la evaluación de ítems para pruebas de evaluación en Matemática y se les enviaron los ítems a valorar.

Se consideraron como ítems aptos los que obtuvieron el rango de puntajes más altos dados por los evaluadores. Si había una gran dispersión entre las valoraciones se apeló a los comentarios para decidir si el ítem se consideraba apto.

También se estableció la proporción de ítems para cada categoría teniendo en cuenta la preponderancia que se les da en los programas curriculares a esos contenidos. Se determinó que cada categoría podía tener una proporción de ítems distribuidos de la siguiente forma.

Tabla 1: Proporción de ítems de Matemática en cada categoría

CATEGORÍA	PROPORCIÓN DE ÍTEMS
Números y Operaciones	15-18%
Álgebra	45-52%
Geometría	15-20%
Análisis de datos, estadística y probabilidad	9-14%

Fuente: Rodríguez (2016).

Finalmente, se eligieron 88 ítems para la prueba de matemática que conformaron dos cuadernillos con 44 ítems cada uno. La proporción final de ítems por categoría así como la tabla de especificaciones se puede leer en Rodríguez (2016). En la fig. 1 se muestra el proceso completo de establecimiento de estándares y creación de la prueba de matemática.

Los estándares de contenido para la prueba de lectura se elaboraron siguiendo las “Pautas de referencia sobre los niveles de lectura en español como primera lengua” del Programa de Lectura y Escritura en Español (ANEP, 2011). Estas pautas establecen categorías parametrizadas que permiten describir los conocimientos y aptitudes lectoras de los estudiantes. La categoría L4B que corresponde con el lector que finalizó el bachillerato, en sus tres categorías: componentes de lectura, conocimiento lingüístico y géneros discursivos es la que se utiliza como referencia. En la figura 2 se puede apreciar el proceso que dio lugar a la prueba de lectura.

También para la prueba de lectura se utilizaron ítems testeados previamente de tres fuentes diferentes: prueba de evaluación diagnóstica en la dimensión Lectura de Rodríguez, Díaz y Correa (2014), ítems liberados de PISA 2009 e ítems elaborados por el Programa de Lectura y Escritura Académica (LEA) de la Comisión Sectorial de Enseñanza de la Universidad de la República y aplicados a la generación de ingreso 2014.

Se estableció que la proporción adecuada de ítems por cada categoría fuera distribuida de la siguiente forma.

Tabla 2. Proporción de ítems de Lectura en cada categoría.

CATEGORÍA	PROPORCIÓN DE ÍTEMS
Componentes de Lectura	60-75%
Conocimiento lingüístico	15-20%
Géneros discursivos	1-5%

Fuente: Rodríguez (2016).

Para la selección de los ítems se utilizó la información disponible acerca de sus propiedades y se procedió a adaptar los ítems que lo requirieran. La prueba de lectura quedó conformada por 37 ítems.

En la siguiente figura 2 se muestra el proceso de creación de la prueba de Lectura.

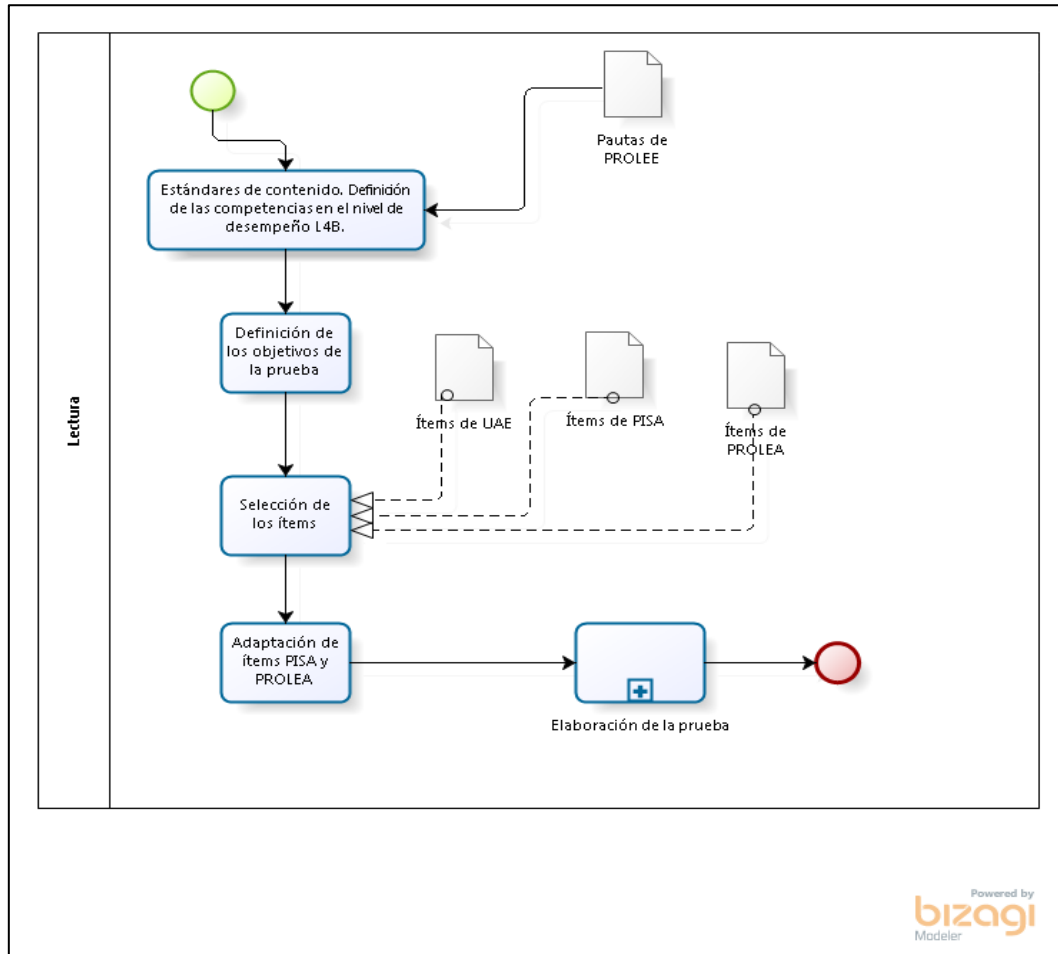


Figura 2. Proceso de creación de la prueba de lectura
Fuente: Elaboración propia.

3.1. Muestra

Se aplicó la prueba de Matemáticas, con sus dos cuadernillos y la prueba de Lectura a los estudiantes que ingresaron a los Centros Universitarios Regionales (CENUR) de la Udelar (Centros Universitarios Regional del Este, Litoral Norte y Noreste, que incluyen las sedes de Maldonado, Paysandú, Rivera, Rocha, Salto, Tacuarembó, Treinta y Tres y Melo) en 2015. La tasa global de respuesta fue del 65,8%. Aunque se apuntaba a un censo, los dos meses que transcurrieron entre la inscripción a la Universidad y la aplicación de la primera prueba hizo que una proporción de los estudiantes hubiera decidido no comenzar los cursos, y por lo tanto, vieron innecesario cumplir con la realización de las pruebas. La evaluación de matemática, realizada en primer lugar, fue cumplimentada por 1.380 estudiantes y la de lectura, aplicada en segunda instancia, fue realizada por 1244 estudiantes. Como se usa TRI no es necesario que la muestra sea representativa por el principio de invarianza.

Un análisis detallado de la cobertura y de las características de la muestra se pueden leer en Rodríguez, Figueroa y Fernández (2016).

Las pruebas se aplicaron a través de la plataforma del Entorno Virtual de Aprendizaje (EVA) de la Udelar. Este espacio proporciona la posibilidad de realizar una prueba fija informatizada.

4. Resultados

Se realizaron una serie de análisis psicométricos a efectos de estudiar las propiedades de las pruebas. Se consideraron para este estudio aquellos estudiantes que no tuviesen más de cinco ítems omitidos.

En primer lugar, se calcularon los índices de dificultad e índices de discriminación de cada ítem así como la distribución de las respuestas.

La dificultad media de los ítems del cuadernillo 1 de matemática fue de 0,36 (0,02), del cuadernillo 2 fue 0,37 (0,02) y en la prueba de lectura se encontró en 0,52 (0,03). Se recomienda que el índice de dificultad sea mayor a 0,5 y que la dificultad media sea 0,62 para ítems de 4 opciones (Abad et al., 2011). Si bien los ítems de nuestras pruebas tienen 5 distractores, nos encontramos alejados de lo recomendado para el índice de dificultad en las pruebas de Matemática y en el rango en la de Lectura. Hay que resaltar que las pruebas de Matemática fueron elaboradas siguiendo estándares que contemplan los contenidos mínimos en Matemática para los egresados de los bachilleratos. Esto nos demuestra que, aunque los ítems evalúan competencias básicas, resultaron difíciles a la muestra en que se aplicó la prueba.

Fueron calculados los índices de discriminación, en este caso las correlaciones biseriales puntuales (bivariadas) de los ítems por separado y de los grupos. La discriminación media de los ítems del cuadernillo 1 y 2 de Matemática fue de 0,40 (0,01 y 0,02 respectivamente) y para la prueba de Lectura se situó en 0,33 (0,01). La bibliografía sugiere que los ítems que obtengan un índice de discriminación menor a 0,20 sean descartados (Abad et al., 2011; Schmeiser y Welch, 2006). Este criterio se usó para la eliminación de los ítems que no ajustaban. En esta primera instancia no se eliminaron ítems del cuadernillo 1 de matemática; se sacaron los ítems 9, 11, 19, 28 y 31 del cuadernillo 2 de Matemática y se eliminaron los ítems 9, 19, 29 y 34 de la prueba de Lectura. Una vez eliminados estos ítems se calculó el coeficiente alfa de Cronbach para determinar la fiabilidad de las pruebas.

El alfa de Cronbach de las pruebas de Matemática fueron 0,88 para el cuadernillo 1 y 0,87 para el cuadernillo 2. El alfa de la prueba de Lectura se situó en 0,75. (Rodríguez et al., 2015). Los coeficientes de fiabilidad son muy buenos para la prueba de Matemática y bueno para la prueba de Lectura.

Para estudiar la validez se analizó la unidimensionalidad de las pruebas. Con la finalidad de conocer si la matriz de datos es adecuada para realizar un análisis factorial se halla el índice Kaiser-Meyer-Olkin y se efectúa el test de esfericidad de Barlett.

El valor global del índice de adecuación muestral de KMO para el cuadernillo 1 de Matemática es de 0,913 y de 0,899 para el cuadernillo 2, que según la clasificación de Kaiser se puede considerar como “maravillosa”. Para la prueba de Lectura el índice de KMO es de 0,836, que se considera “meritoria”.

En el test de esfericidad de Barlett el valor del chi-cuadrado para el cuadernillo 1 de Matemática es 5524,307(946) p-value=0, para el cuadernillo 2 de Matemática es

4643,2(946) p-value=0 y para el cuadernillo de Lectura es 3697,972(666) p-value=0. Por lo tanto, en los tres casos podemos rechazar la hipótesis nula de que las variables utilizadas en el análisis no se correlacionarían en la población en la que se ha extraído la muestra.

A partir de los dos estudios anteriores podemos concluir que los datos son adecuados para realizar un análisis factorial. Se realiza un análisis de componentes principales con rotación varimax para determinar si las dimensiones a priori son unidimensionales.

Con el propósito de evaluar la unidimensionalidad se utilizó el criterio propuesto por Reckase (1979) donde se pide que el primer componente explique, por lo menos, el 20 por ciento de la varianza. El porcentaje de varianza explicada por el primer factor es 26,37% y el ratio entre los dos primeros valores propios es 3,85 para el cuadernillo 1 de Matemática. Para el cuadernillo 2 de Matemática la varianza explicada por el primer factor es 26,69% y el ratio entre los dos primeros valores propios es 4,92. En la prueba de Lectura la varianza explicada por el primer factor es 18,46% y el ratio entre los dos primeros valores propios es 3,98. Si bien en este último caso la varianza del primer factor no alcanza al 20%, se destaca que el ratio entre los dos primeros valores propios es alto.

También se utilizó el cociente entre el primer autovalor y el segundo. Este cociente será indicativo de unidimensionalidad si es aproximadamente 4. En el cuadernillo 1 de Matemática el cociente entre el primer autovalor y el segundo es de 3,84, en el cuadernillo 2 de Matemática es de 4,91 y en el cuadernillo de Lectura es de 3,98.

A partir de estos análisis realizados se puede considerar que las pruebas son unidimensionales y, por lo tanto, es posible aplicar la TRI.

En primer lugar, se calculan las curvas características de los ítems (CCI) y los parámetros de dificultad y discriminación de cada cuadernillo de Matemática y de la prueba de Lectura.

Para calibrar nuestra prueba nos basamos en la TRI y se utilizó el modelo logístico de 2 parámetros. Se estudió el ajuste al modelo tanto de las curvas características del ítem (CCI) como de la habilidad de las personas (*person fit*). Para medir el ajuste de las personas usamos el estadístico L_o de Levine y Rubin (1979) y su versión estandarizada L_z propuesta por Drasgow et al. (1985). A través de este análisis se encontró que el 96,45 % de los sujetos ajustaron bien en el cuadernillo 1 de Matemática, el 98,46 % ajustaron bien para el cuadernillo 2 y el 96,44% para el cuadernillo de Lectura (Rodríguez, 2016).

Se equipararon los cuadernillos 1 y 2 de Matemática mediante ítems de anclaje usando el método de media y desviación (Hambleton, Swaminathan y Rogers, 1991). Los coeficientes para llevar los ítems de la escala del cuadernillo 1 a la escala común fueron $\alpha = 0,91$ y $\beta = 0,13$. Para el cuadernillo 2 fueron $\alpha = 1,08$ y $\beta = -0,05$.

4.1. Establecimiento del punto de corte

La elección de un método para el establecimiento del punto de corte es un tema crucial en el establecimiento de estándares y en la interpretación de los resultados (Jornet, González Such y Suárez, 2010). Con el objetivo de lograr la mayor independencia entre la dificultad empírica y los puntos de corte se optó por el método propuesto por García

et al. (2013), ya que los ítems son diseñados, o en nuestro caso, seleccionados en base a los estándares de desempeño establecidos para clasificar a los estudiantes.

Este nuevo método, basado en la TRI, consta de cinco pasos que se sintetizan a continuación:

1. Construcción o selección de un banco de ítems basado en los estándares de contenido.
2. Calibración del banco de ítems y estimación de las curvas características de los ítems (CCI). En nuestro caso se utilizó el modelo logístico de 2 parámetros.
3. Cálculo de las CCI promedio para cada familia de ítems (para todos los ítems que se encuentran en el mismo nivel de desempeño).
4. Cálculo de las CCI promedio conjuntas para cada familia de ítems.
5. Cálculo del punto de corte. Para nuestras pruebas se calcularon 3 puntos de corte de forma de clasificar a los sujetos en los diferentes niveles de desempeño.

Los estudiantes quedaron clasificados según el θ obtenido de la siguiente forma.

Tabla 3. Clasificación de los estudiantes según la habilidad obtenida en las pruebas de Matemática

θ	NIVEL
<0.62	Insuficiente
≥ 0.62 y <2.06	Suficiente
≥ 2.06	Avanzado

Fuente: Rodríguez (2016).

4.2. Desempeño en Matemática

El 78,06% de los estudiantes quedaron en el nivel Insuficiente, el 18,3% en el nivel Suficiente y el 3,65% en el nivel Avanzado. Estos resultados son coherentes con los obtenidos a través de otra prueba de evaluación diagnóstica en Matemática para estudiantes de ingreso a la Universidad, esto es, distintos ítems y diferente método para el cálculo del punto de corte. En ese caso se utilizó una aplicación del método de Angoff para clasificar a los estudiantes en dos niveles (suficiente e insuficiente). Se compararon los resultados con la dimensión Resolución de problemas, que evaluó competencias matemáticas para el nivel de egresados de bachillerato (Rodríguez, Díaz y Correa, 2014). La proporción de suficientes estuvo en el 22%. Si sumamos el nivel de desempeño suficiente y avanzado de los resultados obtenidos en 2015 con los instrumentos creados, obtenemos un 21,95% de estudiantes que superaron la suficiencia. También los resultados hallados en otras pruebas de matemática para estudiantes de ingreso a la Universidad presentan proporciones de suficiencia igualmente bajas (Míguez, Blasina y Loureiro, 2013; Mussio y Martinotti, 2013; UEFI, 2012).

Las competencias asociadas al nivel de desempeño Insuficiente indican que el 78,06% son capaces de resolver problemas de aritmética incluyendo porcentajes, razones y proporciones, pueden operar con fracciones, sustituyen y simplifican expresiones algebraicas simples, pueden determinar el límite de una función dada por su gráfica. La mayoría de estas competencias se adquieren durante el primer ciclo de la Educación Media. Algunas de las competencias que se incluyen dentro del nivel Suficiente, alcanzado por el 18,3% de los estudiantes, muestran que son capaces de operar con fracciones usando paréntesis, operar con números complejos y representarlos en el plano

complejo, resolver ecuaciones y desigualdades complejas de una variable real, resolver problemas de conteo utilizando números combinatorios, operar con exponenciales, logaritmos y potencias, calcular la media, mediana, moda, cuartiles, varianza o resolver problemas usando las propiedades de la probabilidad. Estas competencias se corresponden con las que se deben adquirir al término del segundo ciclo de la Educación Media en las orientaciones de bachillerato con menos contenidos de Matemática. La definición completa de las competencias en matemática en cada nivel de desempeño se encuentra en Rodríguez (2016).

4.3. Desempeño en Lectura

Se obtuvieron los θ para clasificar a los estudiantes en los distintos niveles de desempeño. Se presenta en la tabla 4.

Tabla 4. Clasificación de los estudiantes según la habilidad obtenida en la prueba de Lectura

θ	NIVEL
<-0.085	Insuficiente
>=-0.085 y <1.85	Suficiente
>=1.85	Avanzado

Fuente: Rodríguez (2016).

El 46,98% de los estudiantes quedó clasificado en el nivel Insuficiente, 51,18% en el Suficiente y el 1,84% en el Avanzado. Si comparamos estos resultados en Lectura con los obtenidos en la dimensión Comprensión y aplicación de la prueba aplicada en 2012, también podemos afirmar que existe coherencia, ya que en esa instancia hubo un 40% de suficientes (Rodríguez, Díaz y Correa, 2014), mientras que en esta prueba un 53% superaron la suficiencia.

Las competencias de los estudiantes que quedaron clasificados en el nivel Insuficiente incluye relacionar al autor y al texto con su contexto histórico y sociocultural, establecer relaciones intratextuales entre los diferentes bloques del texto, reconocer las estructuras sintácticas del español. Estas competencias están asociadas a los estudiantes que culminaron el primer ciclo de la Educación Media. El 51,18% de los estudiantes que alcanzaron el nivel de Suficiente poseen una lectura reflexiva y crítica, reconocen posibles inconsistencias internas de los textos, reconocen el léxico básico de las disciplinas específicas asociadas con la educación formal. Las competencias de este nivel se asocian a estudiantes que culminaron el segundo ciclo de la Educación Media. La descripción completa de cada nivel de lector se encuentra en ANEP (2011).

5. Discusión

En primer lugar, debemos resaltar el desarrollo metodológico de las pruebas donde se siguió lo recomendado por la literatura especializada para el establecimiento de estándares de contenido y desempeño (Cizek y Bunch, 2007; Jornet, González Such y Suárez, 2010; Tourón, 2009).

Se obtuvieron dos pruebas calibradas mediante TRI y validadas para la evaluación de competencias en matemática y lectura para el nivel de ingreso a la universidad. La validez de contenido de los instrumentos está fundamentada en el procedimiento utilizado para la delimitación de los contenidos a evaluar, es decir, el propio

establecimiento de estándares de contenido. La validez de constructo, obtenida a través de los análisis factoriales, se aprecia en la concordancia entre el contenido de los estándares y los factores obtenidos.

Dentro de los aspectos metodológicos se destaca la utilización de un método para el establecimiento de los puntos de cortes basados en la TRI que aporta solidez al análisis y a la interpretación de los resultados (Muñiz, 1997; Jornet, González Such y Suárez, 2010).

La baja proporción de estudiantes que superaron el nivel de suficiencia es congruente con otras evaluaciones y pone evidencia la vulnerabilidad académica de los estudiantes de los CENUR, que captan mayoritariamente estudiantes locales, es decir, que viven en la capital departamental donde se ubica la Sede, o de la región, provenientes de los pueblos cercanos. La comparación del desempeño en Matemática entre estudiantes de ingreso provenientes de Montevideo contra los del Interior realizada por Mussio y Martinotti (2013) apunta en el mismo sentido, ya que los estudiantes que realizaron la educación preuniversitaria en el Interior obtuvieron menores puntajes en la prueba, es más, cuanto más lejos de la capital se encontraba la residencia del estudiante menor era el puntaje obtenido. Los bajos resultados, tanto en Matemática como Lectura, se corresponden con los hallazgos de la investigación sobre “Trayectorias académicas y laborales de los jóvenes uruguayos - El panel PISA 2003-2007” de Boado y Fernández (2010). Los estudiantes con menores puntajes que lograron acceder a la educación superior fueron los que cursaron educación media en el Interior del país.

Por otra parte, la devolución de resultados a los estudiantes, obteniendo una retroalimentación inmediata sobre sus aciertos y errores, además de un comentario global con sugerencias según el nivel de desempeño que se encontraban, les permitió elegir trayectorias alternativas o tomar cursos compensatorios que se crearon especialmente para apoyar las dificultades detectadas. Las investigaciones realizadas en estudiantes con baja preparación en matemática para afrontar las exigencias universitarias indican que los cursos de apoyo o de nivelación mejoran los resultados (Bettinger y Long, 2009; Hillock et al., 2013; Perkin y Bamforth, 2009).

6. Conclusiones

Sobre los resultados en las pruebas se puede concluir que, teniendo en cuenta los estándares de contenido y la complejidad de los ítems, las proporciones de estudiantes que lograron la suficiencia son muy bajas en Matemáticas (22%) y bajas en Lectura (53%). La coherencia de estos resultados con los obtenidos en otras pruebas con objetivos similares proporciona evidencia sobre la solidez de los instrumentos (Mussio y Martinotti, 2013; Rodríguez, Díaz y Correa, 2013, 2014; UEFI, 2012).

Una mención especial merece la definición de políticas educativas en función de los resultados obtenidos, ya que era uno de los objetivos planteados. Se implementaron dos programas compensatorios en el área de Matemática, uno en la región Noreste y otro en el CENUR Este, como forma de apoyo para los estudiantes que fueron clasificados en el nivel Insuficiente. Por otra parte, el Programa de Lectura y Escritura Académica de la Comisión Sectorial de Enseñanza impartió talleres de apoyo en lectura para todos los estudiantes que lo requirieran. Algunas de las primeras políticas implementadas se describen en Rodríguez, Figueroa y Fernández (2016).

Además, desde la Comisión Coordinadora del Interior se han impulsado varias líneas de acción: se institucionalizó la Evaluación Diagnóstica como un programa y su Plenario resolvió continuar la aplicación de las pruebas para 2016; se resolvió integrar un equipo académico en Lectura y Escritura para elaborar un programa de apoyo en Lectura (CCI, 2016a); se impulsó trabajo conjunto con la Comisión Sectorial de Enseñanza (CCI, 2016b) y se conformó un Grupo Académico de Matemática con el doble objetivo de elaborar un programa de apoyo en Matemática y colaborar con la creación y evaluación de ítems para la prueba diagnóstica (CCI, 2016c). Este grupo produjo un programa de apoyo en Matemática con características innovadoras, que es recibido por el Plenario de la CCI, encomendando a los coordinadores y comisiones de carrera del Interior estudiar la propuesta para su implementación (CCI, 2016d).

Referencias

- Abad, F., Olea, J., Ponsoda, V., García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.
- Altmark, S., Castrillejo, A., Debera, L. y Nalbarte, L. (2006). *Elaboración de pruebas diagnósticas al ingreso a la Facultad de Ciencias Económicas y Administración DT (06/02)*. Recuperado de <http://www.iesta.edu.uy/wp-content/uploads/2010/03/0602.pdf>
- ANEP. (2011). *Pautas de referencias sobre niveles de lectura en español como primera lengua*. Montevideo: PROLEE, Administración Nacional de Educación Pública.
- Attewell, P., Heil, S. y Reisel, L. (2012). What is academic momentum? And does it matter? *Educational Evaluation and Policy Analysis*, 34(1), 27-44. doi:10.3102/0162373711421958
- Bertoni, E. (2005). *El estudiante universitario: Una aproximación al perfil de ingreso. Documento de trabajo No. 3*. Montevideo: Unidad Académica, Comisión Sectorial de Enseñanza.
- Bettinger, E. P. y Long, B. T. (2009). Addressing the needs of under-prepared students in higher education: Does college remediation work? *The Journal of Human Resources*, 44, 736-771. doi:10.1353/jhr.2009.0033
- Black, P. y Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Boado, M. y Fernández, T. (2010). *Trayectorias académicas y laborales de los jóvenes en Uruguay. El panel PISA 2003-2007*. Montevideo: FCS-UDELAR.
- Burton, N. W. (1978). Societal standards. *Journal of Educational Measurement*, 15(4), 263-271. doi:10.1111/j.1745-3984.1978.tb00073.x
- Cizek, G. J. y Bunch, M. B. (2007). *Standard setting. A guide to establishing and evaluating performance standards on tests*. Thousand Oak, CA: Sage Publications.
- Cizek, G. J., Bunch, M. B. y Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23(4), 31-31. doi:10.1111/j.1745-3992.2004.tb00166.x
- CollegeBoard. (2014). *SAT Practice Questions Math*. Recuperado de <http://sat.collegeboard.org/practice/>
- Comisión Coordinadora del Interior. (2016a). *Resolución No. 28 del 10/02/2016*. Recuperado de <http://www.expe.edu.uy/expe/resoluci.nsf/ed0853334a4fde3c83257c8d007f3233/426a58292b473fb503257f73004431bd?OpenDocument>

- Comisión Coordinadora del Interior. (2016b). *Resolución No. 704 del 29/08/2016*. Recuperado de <http://www.expe.edu.uy/expe/resoluci.nsf/ed0853334a4fde3c83257c8d007f3233/5696d1b256ae8a15032580360052db19?OpenDocument>
- Comisión Coordinadora del Interior. (2016c). *Resoluciones 132 y 133 del 07/03/2016*. Recuperado de <http://www.expe.edu.uy/expe/resoluci.nsf/ed0853334a4fde3c83257c8d007f3233/a586a2373f0d542b03257f9600608060?OpenDocument>
- Comisión Coordinadora del Interior. (2016d). *Resolución No. 1069 del 12/12/2016*. Recuperado de <http://www.expe.edu.uy/expe/resoluci.nsf/ed0853334a4fde3c83257c8d007f3233/b1d92c14f5bf21ea0325808b006734eb?OpenDocument>
- Drasgow, F., Levine, M. V., Williams, B., McLaughlin, M. E. y Candell, G. L. (1989). Modeling incorrect responses to multiple-choice items with multilinear formula score theory. *Applied Psychological Measurement, 13*(3), 285-299. doi:10.1177/014662168901300309
- Enrich, H., Míguez, M., Rodríguez Ayán, M. N. y Leymonié, J. (2006). *Evaluación diagnóstica de las habilidades matemáticas al ingreso en las facultades del área científico-tecnológica*. Montevideo: CSE, Udelar.
- Ferrer, G. (2006). *Estándares en educación. Implicancias en América Latina*. Santiago de Chile: PREAL.
- Ferrer, J. G., Valverde, G. y Esquivel, J. M. (2006). *Aspectos del currículo prescrito en América Latina: Revisión de tendencias contemporáneas en currículo, indicadores de logro, estándares y otros instrumentos. Informe de trabajo*. Chile: PREAL.
- García, P. E., Abad, F. J., Olea, J. y Aguado, D. (2013). A new IRT-based standard setting method: Application to elath-listening. *Psicothema, 25*(2), 238-244.
- Gil Pascual, J. A. (2016). *Técnicas e instrumentos para la recogida de información*. Madrid: UNED.
- Gitomer, D. H., Steinber, L. S. y Mislevy, R. J. (2009). Diagnostic assessment and troubleshooting skill in an intelligent tutoring system. En P. D. Nichols, S. F. Chipman y R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 73-102). Mahwah, NJ: Lawrence Erlbaum Associates.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American psychologist, 18*(8), 519-521. doi:10.1037/h0049294
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*(4), 237-261. doi:10.1111/j.1745-3984.1978.tb00072.x
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. En G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89-116). Nueva York, NY: Lawrence Erlbaum.
- Hambleton, R., Swaminathan, H. y Rogers, H. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks, CA: Sage Publications.
- Hamilton, L. S., Stecher, B. M. y Yuan, K. (2008). *Standards-based reform in the United States: History, research and future directions*. Los Ángeles, CA: RAND Corporation.
- Hillock, P. W., Jennings, M., Roberts, A. y Scharaschkin, V. (2013). A Mathematics support programme for first-year engineering students. *International Journal of Mathematical Education in Science and Technology, 44*(7), 1030-1044. doi:10.1080/0020739x.2013.823251
- Jornet, J. y González Such, J. (2009). Evaluación criterial: Determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación, 16*, 103-121.

- Jornet, J., González Such, J. y Suárez, J. (2010). Validación de los procesos de determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, 19, 11-29.
- Levine, M. V. y Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational and Behavioral Statistics*, 4(4), 269-290. doi:10.2307/1164595
- Lewis, D. M. y Green, D. R. (junio, 1997). The validity of performance level descriptors. Comunicación presentada en el *Annual CCSSO Conference on Large Scale Assessment*, Universidad de Colorado, Springs, CO.
- Linn, R. (2003). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives*, 11(31). Recuperado de <http://epaa.asu.edu/epaa/v11n31/>
- Linn, R. L. y Gronlund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice-Hall.
- Míguez, M., Blasina, L. y Loureiro, S. (septiembre, 2013). *Diagnóstico al ingreso en la Facultad de Ingeniería de la Universidad de la República: Matemática y variables no tradicionales*. Comunicación presentada en el VII Congreso Iberoamericano de Educación Matemática, Montevideo. Recuperado de <http://www.cibem7.semur.edu.uy/7/actas/pdfs/873.pdf>
- Mills, C. N. y Jaeger, R. M. (1988). Creating descriptions of desired student achievement when setting performance standards. En L. Hansche (Ed.), *Handbook for the development of performance standards* (pp. 73-86). Washington D. C.: Council of Chief State School Officers.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Muñiz, J. (1998). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Mussio, I. y Martinotti, L. (2013). *Informe sobre prueba diagnóstica aplicada a estudiantes que ingresan a la Facultad de Ciencias Sociales*. Recuperado de <http://cienciassociales.edu.uy/departamentodeeconomia/wp-content/uploads/sites/2/2013/archivos/1213.pdf>
- O'Shea, M. (2005). *From standards to success*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Pérez Juste, R. (2006). *Evaluación de programas educativos*. Madrid: La Muralla.
- Perkin, G. y Bamforth, S. (2011). A variety of approaches to the provision of mathematics help for first-year engineering undergraduates. *International Journal of Electrical Engineering Education*, 48(1), 80-91. doi:10.7227/ijeee.48.1.7
- Prieto, G. y Delgado, A. (1996). Construcción de ítems. En J. Muñiz (Coord.), *Psicometría* (pp. 105-138). Madrid: Universitas.
- Ravitch, D. (1996). *Estándares Nacionales en Educación*. Santiago de Chile: PREAL.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*, 4(3), 207-230. doi:10.2307/1164671
- Rodríguez, P. (2014). Oportunidades y riesgos en el acceso a la Educación Superior en el marco del Centro Universitario de la Región Este. En T. Fernández y A. Ríos (Eds.), *El tránsito entre ciclos en el Educación Media y Superior de Uruguay* (pp. 165-181). Montevideo: CSIC.
- Rodríguez, P. (2016). *Creación y establecimiento de estándares para la evaluación de la calidad de la educación superior: Un modelo adaptado a los centros universitarios de la Udelar* (Tesis doctoral). Universidad Nacional de Educación a Distancia, Madrid.

- Rodríguez, P. y Correa, A. (2011). *Informe de las Evaluaciones Estudiantiles 2011. Evaluaciones de centro, docente y asignaturas*. Maldonado: CURE.
- Rodríguez, P., Correa, A. y Díaz, M. (2012). *Informe sobre Evaluación Diagnóstica 2012*. Maldonado: CURE.
- Rodríguez, P., Díaz, M. y Correa, A. (2013). *Resultados de la Evaluación Diagnóstica 2013*. Maldonado: CURE.
- Rodríguez, P., Díaz, M. y Correa, A. (2014). Los aprendizajes al ingreso en un Centro Universitario Regional. *Intercambios*, 2(1), 91-100.
- Rodríguez, P., Carreño, G., Fernández, T., Figueroa, V. y Lorda, N. (2015). *Evaluación diagnóstica 2015 en Matemática y Lectura. Primer informe de resultados*. Montevideo: CCI.
- Rodríguez, P., Figueroa, V. y Fernández, T. (2016). Evaluación de competencias al ingreso a la Universidad. *Temas de Educación*, 22(1).
- Rupp, A., Templin, J. y Henson, R. (2010). *Diagnostic measurement. Theory, methods and applications*. Nueva York, NY: The Guilford Press.
- Schmeiser, C. B. y Welch, C. J. (2006). Test development. En R. L. Brennan (Ed.), *Educational measurement* (pp. 307-353). Westport, CT: Praeger Publishers.
- Sotomayor, C. y Gysling, J. (2011). Estándares y regulación de calidad de la formación de profesores: Discusión del caso chileno desde una perspectiva comparada. *Calidad en la Educación*. (35), 91-129. doi:10.4067/s0718-45652011000200004
- Tourón, J. (2009). El establecimiento de estándares de rendimiento en los sistemas educativos. *Estudios sobre Educación*, 16, 127-146.
- Unidad de Enseñanza de Facultad de Ciencias. (2005). *Evaluación Diagnóstica de Conocimientos y Habilidades al Ingreso*. Recuperado de <http://ue.fcien.edu.uy/Documentos/Edich2005.pdf>
- Unidad de Enseñanza de Facultad de Ciencias. (2010). *Evaluación Diagnóstica de Conocimientos y Habilidades al Ingreso*. Recuperado de http://ue.fcien.edu.uy/Documentos/EDICH_2010_primer_informe.pdf
- Unidad de Enseñanza de Facultad de Ingeniería. (UEFI). (2012). *Informe de la Herramienta Diagnóstica al Ingreso*. Recuperado de <https://www.fing.edu.uy/sites/default/files/noticias/2012/7/5793/Informe%20final%20%20HDI2012.pdf>
- Zalba, M. E., Gómez de Erice, M. V., Alfonso, V., Deamici, C., Erice, X., Gutiérrez, N. B., ... Sayavedra, C. (2005). *Competencias para el ingreso y permanencia en la universidad: Una propuesta de articulación curricular entre el nivel superior y el nivel medio de enseñanza: La experiencia de la Universidad Nacional de Cuyo*. Recuperado de <http://www.cinda.cl/download/libros/Curr%C3%ADculo%20Universitario%20Basado%20en%20Competencias.pdf>

Breve CV de la autora

Pilar Rodríguez Morales

Doctora en Educación por la UNED, España. Profesora Adjunta, Coordinadora de la Unidad de Apoyo a la Enseñanza del Centro Universitario Regional del Este, Universidad de la República de Uruguay. Corresponsable del Programa de Evaluación

Diagnóstica de la Comisión Coordinadora del Interior de la Universidad de la República.
Integrante del Sistema Nacional de Investigadores desde 2016. ORCID ID: 0000-0003-
1929-4961. Email: prodriguez@cure.edu.uy