



Universidad de Granada
Facultad de Ciencias de la Educación
Departamento de Didáctica de la Matemática

ANÁLISIS DE LOS ÍTEMS DE LAS PRUEBAS DE EVALUACIÓN DE
DIAGNÓSTICO EN COMPETENCIA MATEMÁTICA PARA EL SEGUNDO CURSO
DE LA EDUCACIÓN SECUNDARIA OBLIGATORIA EN
ESPAÑA, 2008-2009: UN ESTUDIO EXPLORATORIO

Rosa Marta Caraballo Caraballo
Granada, 2010



Universidad de Granada
Facultad de Ciencias de la Educación
Departamento de Didáctica de la Matemática

ANÁLISIS DE LOS ÍTEMS DE LAS PRUEBAS DE EVALUACIÓN DE DIAGNÓSTICO
EN COMPETENCIA MATEMÁTICA PARA EL SEGUNDO CURSO
DE LA EDUCACIÓN SECUNDARIA OBLIGATORIA EN ESPAÑA, 2008-2009:
UN ESTUDIO EXPLORATORIO

Trabajo de Investigación Tutelada presentado por
D^a. Rosa Marta Carballo Carballo para optar por el máster en
Didáctica de la Matemática, bajo la dirección del Dr. D. Luis Rico Romero.

Fdo.: Rosa Marta Carballo Carballo

Vo. Bo. del Director

Fdo. Dr. Luis Rico Romero

Granada, 2010

Este estudio fue realizado dentro del grupo de investigación *Didáctica de la Matemática. Pensamiento Numérico* de la Universidad de Granada, perteneciente al Plan Andaluz de Investigación, Desarrollo e Innovación de la Junta de Andalucía (FQM193) en las líneas de investigación *Diseño, Desarrollo e Innovación del Currículo de Matemáticas y Calidad y Evaluación de Planes de Formación en Matemáticas*.

DEDICATORIA

A...

Alexis, compañero de travesía.

Mónica, Natalia y Gabriel, razones para caminar.

Doña Juanita, por iniciar mi camino.

Don Ernesto, por siempre alumbrar mi senda aún desde su ausencia eterna.

AGRADECIMIENTOS

Dios quiso que este último año formara parte de mi travesía. Comenzar a estudiar un posgrado y completar un máster en Didáctica de la Matemática en la Universidad de Granada ha enriquecido mi vida de muchas maneras. Con este proceso he crecido profesional y personalmente y he creado lazos de amistad por los cuales me siento sumamente agradecida.

Quiero agradecer a todas las personas que han hecho posible que este trabajo viera finalmente la luz. Agradezco al profesor Dr. Luis Rico Romero por su apoyo, su tiempo, sus consejos y sus críticas edificantes, continuas e invaluable que lograron guiarme en la dirección correcta; sus sugerencias, alternativas e inspiración demostraron ser de utilidad para mis luchas de cada día. Mi gratitud especial al Dr. Moisés Coriat por su tiempo y su estímulo para que hallara el valor de solicitarle al Dr. Rico que dirigiera esta investigación. Gracias al Dr. Enrique Castro por su valioso apoyo técnico. Al Dr. José Luis Lupiáñez por su tiempo y sus útiles y valiosas sugerencias. A mis “amigos de la locura” (ustedes saben quienes son) por prestarme sus oídos y hombros y brindarme aliento. Sin ustedes, seguir el rumbo y conservar la cordura hubiera sido misión imposible. A los miembros de la facultad del departamento de Didáctica de la Matemática y a mis compañeros por su continua solidaridad.

Finalmente, y realmente lo más importante, mi más profundo agradecimiento a mis seres más amados: mi esposo, mis hijos, mi padre (QEPD), mi madre y mis hermanos. Gracias por su amor constante, fe ciega y apoyo incondicional a la realización de este sueño largamente acariciado. Gracias por ser mis más fervientes admiradores y por estar siempre en el lugar exacto, donde más los he necesitado. Siempre detrás del telón, ustedes han hecho posible mi protagonismo.

Alguna vez leí que lo importante no es el destino sino la travesía; he caminado un tramo maravilloso del camino, es tiempo de comenzar a andar de nuevo.

Rosa Marta Caraballo Caraballo
Granada, España
Septiembre, 2010

Índice

Capítulo 1

INTRODUCCIÓN AL ESTUDIO	1
1. Aproximación al problema.....	1
2. Justificación del estudio.....	4
3. Objetivos del estudio	6
4. Estructura del documento.....	7

Capítulo 2

MARCO DE REFERENCIA DE LA INVESTIGACIÓN	8
1. Ley Orgánica de Educación (LOE).....	8
2. Las evaluaciones generales y las evaluaciones autonómicas de diagnóstico.....	9
2.1. Condiciones técnicas de aplicación.....	12
3. El modelo matemático de PISA	13
3.1. Caracterización de las pruebas PISA.....	16
3.1.1. Variables de tarea	17
3.1.2. Variables de sujeto	19
3.1.3. Variables de resultado o de desempeño.....	19
4. Las variables de tarea de PISA como criterio para evaluar los ítems	23
5. El enfoque funcional del Aprendizaje Matemático.....	25
6. La Teoría de Respuesta a los Ítems y el modelo de Rasch	26

Capítulo 3

METODOLOGÍA DE INVESTIGACIÓN	29
1. Búsqueda y recopilación de las pruebas de diagnóstico	29
2. Definición de la muestra	30
3. Tipo de estudio	31
4. Técnicas de análisis.....	32
5. Limitaciones del estudio	32

Capítulo 4

ANÁLISIS DE LOS ÍTEMS	33
1. Análisis de los ítems por Comunidad Autónoma.....	33
2. Análisis de los ítems por variables cruzadas.....	39
3. Análisis logarítmico- lineal: Interacción entre las variables	43
4. Evaluación de los ítems según las variables de tarea	44
4.1 Ítem en el contexto personal	45
4.2 Ítem en el contexto educativo/laboral.....	46
4.3 Ítem en el contexto público	47
4.4 Ítem en el contexto científico.....	47
5. Características comunes de los ítems	50
5.1 Presentación de la situación.....	50
5.2 Fortalezas y deficiencias de los ítems	51
5.2.1 Ejemplos de ítems con fortalezas	55
5.2.2 Ejemplos de ítems con deficiencias	57

Capítulo 5

EVALUACIÓN DE LOS RESULTADOS Y CONCLUSIONES	61
1. Resumen de los hallazgos	61
2. Evaluación de los resultados	63
2.1 Evaluación del objetivo 1	63
2.2 Evaluación del objetivo 2.....	64
2.3 Evaluación del objetivo 3	64
2.4 Evaluación del objetivo general.....	64
3. Conclusiones, implicaciones y recomendaciones.....	65

Referencias	67
--------------------------	----

Índice de anexos	69
-------------------------------	----

ÍNDICE DE TABLAS

1. Distribución de los ítems por Comunidad Autónoma	31
2. Distribución de los ítems por Comunidad Autónoma y contexto	34
3. Porcentaje de los ítems por Comunidad Autónoma y contexto.....	34
4. Distribución de los ítems por Comunidad Autónoma y contenido	36
5. Porcentaje de los ítems por Comunidad Autónoma y contenido.....	36
6. Distribución de los ítems por Comunidad Autónoma y nivel de complejidad..	38
7. Porcentaje de los ítems por Comunidad Autónoma y nivel de complejidad.....	38
8. Porcentaje de los ítems por contexto, complejidad y contenido.....	40
9. Porcentaje de los ítems por contenido y nivel de complejidad.....	41
10. Porcentaje de los ítems por contexto y nivel de complejidad.....	42
11. Porcentaje de los ítems por contenido y contexto	43
12. Resultado de las pruebas estadísticas de asociaciones parciales	44
13. Análisis de ítems en el contexto personal.....	45
14. Análisis de ítem en el contexto educativo/laboral	46
15. Análisis de ítem en el contexto público.....	48
16. Análisis de ítem en el contexto científico.....	50
17. Distribución de los ítems por la forma de presentar la situación.....	51
18. Porcentaje de ítems con fortalezas.....	53
19. Porcentaje de ítems con debilidades	54
20. Capacidades que potencialmente movilizan la tareas.....	55

ÍNDICE DE FIGURAS

1. Funciones de la evaluación	2
2. El ciclo de la matematización.	16
3. Disponibilidad de las pruebas por Comunidad Autónoma	30
4. Porcentaje de los ítems por contexto y Comunidad Autónoma	35
5. Porcentaje de los ítems por contenido y Comunidad Autónoma	37
6. Porcentaje de los ítems por nivel de complejidad y Comunidad Autónoma	39

CAPÍTULO 1

INTRODUCCIÓN AL ESTUDIO

Este documento es un estudio sobre las pruebas elaboradas para la aplicación de la primera evaluación de diagnóstico en competencia matemática realizada en España por las Consejerías de Educación de las Comunidades Autónomas a los estudiantes de segundo curso de la Educación Secundaria Obligatoria (ESO) en el año académico 2008-2009. Presentamos en este capítulo una aproximación al problema que se estudia, los objetivos que proveen orientación a la investigación, la justificación del estudio y una descripción de la estructura del documento.

1. Aproximación al problema

Evaluar es una tarea necesaria e ineludible. Boulmetis y Dutwin (2000) definen la evaluación como “el proceso sistemático de recopilar y analizar datos con el propósito de determinar si, y hasta qué grado, se han logrado los objetivos propuestos” (p. 4). Un proceso evaluativo con un seguimiento adecuado no sólo provee datos para diagnosticar sobre los elementos participantes en una intervención particular sino que usa este diagnóstico para mejorar las actividades educativas subsiguientes para otros participantes. Evaluar es, de manera sucinta, recoger y sistematizar información para tomar decisiones.

Maestro (2005) identifica cuatro funciones de la evaluación: rendir cuentas, informar, comparar y mejorar (Figura 1). Es indispensable la rendición de cuentas a la sociedad para que sus miembros sientan satisfacción por el esfuerzo y los recursos que invierten en la educación como servicio público. De igual manera, la evaluación provee la información necesaria para adoptar cambios en las políticas educativas que redundarán en el mejoramiento de la calidad y el éxito de los alumnos.

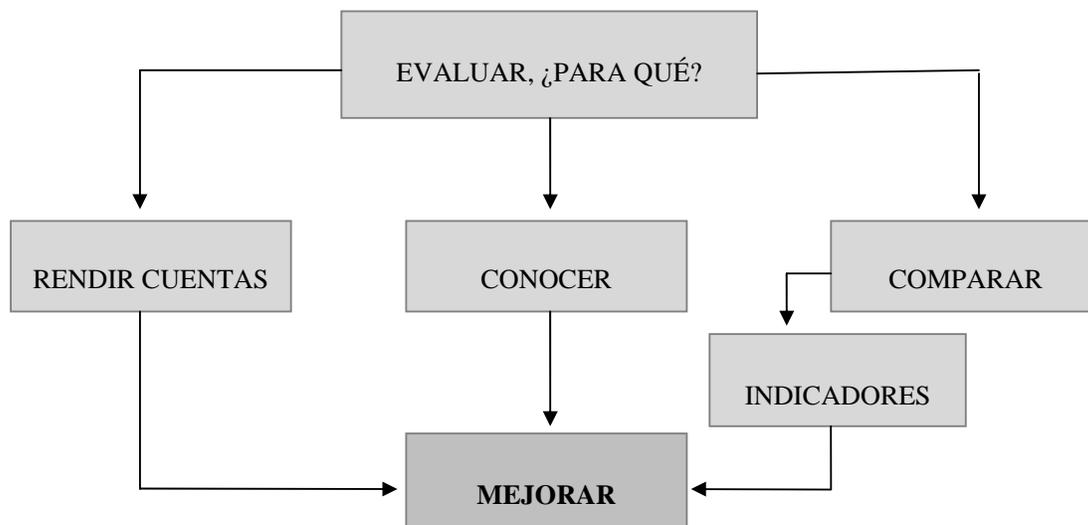


Figura 1. Funciones de la evaluación

No obstante, no es suficiente con conocer los resultados de la evaluación en un momento determinado, es necesario que la evaluación se convierta en un recurso continuo y permanente que permita conocer la realidad del sistema a lo largo del tiempo. Esta saludable práctica permitirá, a su vez, establecer comparaciones no sólo dentro del propio sistema sino con otros sistemas locales, regionales e internacionales. De esta manera los resultados de la evaluación se convierten en indicadores útiles para examinar la realidad educativa y su evolución. En síntesis, el objetivo final de la evaluación es el mejoramiento de la calidad del sistema y, con ello, el compromiso de todos los involucrados para mantenerla.

Por lo tanto, en un sistema educativo, los profesores, administradores, directores y gerentes educativos se benefician de los procesos de evaluación que, como parte de su labor, realizan. Los sistemas de evaluación a nivel nacional se establecen con el propósito principal de mejorar la calidad de las estructuras educativas. Tradicionalmente ha existido una demanda continua por que se produzcan cambios e innovaciones que redunden en una educación más relevante y de mejor calidad. Esta demanda impone la evaluación como condición necesaria para el logro de tal propósito. Incorporar la evaluación a los procesos educativos constituye un requisito de la planificación que conduce a su transformación (Boulmetis & Dutwin, 2000).

La evaluación educativa persigue valorar los méritos y el impacto de los distintos programas e iniciativas que se implementan. Los resultados de una evaluación les permite a los educadores y administradores determinar el éxito de sus programas, identificar esfuerzos y diseñar planes de acción para mejorarlos. Las comunidades educativas que, conscientes de su importancia, han adoptado una cultura de evaluación poseen una herramienta poderosa para identificar y reflexionar respecto de las fortalezas y debilidades de sus prácticas educativas y para tomar decisiones que redunden en el mejoramiento del sistema y, por consiguiente, en el éxito de sus estudiantes.

La Ley Orgánica 2/2006, del 3 de mayo, de Educación (LOE) estipula en sus artículos 21 y 29 que las Administraciones Educativas de cada Comunidad Autónoma realizarán una evaluación de diagnóstico de carácter externo, censal y formativo a todos los alumnos que finalicen el segundo ciclo de la Educación Primaria y el segundo curso de la ESO, respectivamente. Se espera que estas evaluaciones generales de diagnóstico contribuyan a mejorar la calidad y la equidad de la educación arrojando luz sobre la situación del sistema educativo mediante la valoración de los aprendizajes de los alumnos y el impulso de procesos de innovación y mejora del sistema educativo en su conjunto. Además de contribuir a mejorar la calidad del sistema, las evaluaciones de diagnóstico persiguen la finalidad de orientar políticas educativas, aumentar la eficacia y la transparencia de la educación como sistema e informar respecto del grado de adquisición y logro de las competencias básicas del currículo tanto a nivel nacional como regional (Artículo 140, LOE 2/2006).

Con el propósito de conocer el trabajo que las Comunidades Autónomas han realizado en el desarrollo, el diseño y aplicación de las primeras pruebas de diagnóstico en competencia matemática, se analizaron los ítems de dichas pruebas aplicados en el año académico 2008-2009, tomando como referencia el modelo de las evaluaciones para matemáticas del *Programme for International Student Assessment* (PISA) de la *Organization for Economic Co-operation and Development* (OECD). Dicha organización, que en lo sucesivo denominaremos OCDE por sus siglas en español, estableció las evaluaciones PISA con el objetivo de determinar, entre otros, el grado de competencia matemática que los jóvenes de 15 años han adquirido para enfrentarse a situaciones de la vida cotidiana. Uno de los objetivos principales de la OCDE es establecer indicadores de calidad para determinar cómo los sistemas educativos logran ese nivel de formación (Rico, 2007). Las evaluaciones

autónomas de diagnóstico usan el modelo de evaluación en matemáticas establecido por el estudio PISA como marco de referencia para diseñar las pruebas (Lupiáñez, 2010).

2. Justificación del estudio

Por el rol preponderante que desempeñan las competencias en la caracterización de las expectativas de aprendizaje, los resultados obtenidos por los estudiantes españoles en las pruebas PISA de 2003 y de 2006 incidieron en las nuevas orientaciones curriculares. Asimismo, tuvieron el efecto de enfatizar en el carácter funcional del aprendizaje de las matemáticas escolares (Lupiáñez, 2009). El enfoque funcional del aprendizaje matemático apoya que los estudiantes en el nivel de Educación Obligatoria, pongan en juego el conocimiento teórico y el dominio técnico que han adquirido para resolver, de una manera eficaz y eficiente, una variedad de situaciones-problemas. Este enfoque ha sido consistentemente avalado por el proyecto PISA de la OCDE. Tanto las recientes orientaciones curriculares en España a nivel nacional como el marco teórico de las evaluaciones de PISA, respaldadas ambas por el enfoque funcional del aprendizaje matemático, sirven de marco de referencia para esta investigación.

El propósito fundamental de la evaluación de diagnóstico es ofrecer información importante sobre la situación del alumnado, los centros y el sistema educativo como un conjunto. Asimismo sirven como un instrumento potencial para evaluar el desarrollo de las competencias básicas de los escolares. Con los resultados obtenidos se espera que las Comunidades Autónomas adopten planes de mejoramiento de la calidad del sistema. Con la mirada puesta en el mejoramiento de la calidad del sistema educativo español, ¿cuál es la importancia de analizar los ítems aplicados en las evaluaciones de diagnóstico de las Comunidades Autónomas? El análisis de los ítems o unidades de evaluación permite conocer la efectividad de la prueba en el logro de su propósito fundamental. Hamill y Usala (2002) señalan que mediante el análisis de los ítems en una prueba de papel y lápiz es posible identificar:

- Errores potenciales en el proceso de corrección de las pruebas
- Ambigüedad en los ítems
- Si la distribución de los ítems respecto de las variables dentro del marco teórico, se hizo de manera equitativa
- Alternativas que no responden a los propósitos de la prueba

- Ítems que inducen a respuestas erróneas
- Ítems con límite de tiempo
- Déficits o limitaciones de los ítems
- Ítems que no cumplen el propósito para el cual fueron diseñados

Apoyado en los trabajos de Bell, Burkhardt y Swan de 1992, Rico (1997) describe algunas condiciones que deben cumplir las tareas de evaluación de los alumnos.

1. Relevancia práctica. Se presentan situaciones de la vida real pero plantean cuestiones que carecen de significado práctico.
2. Coherencia o fragmentación de la tarea. La tarea conduce al estudiante a través de una secuencia de pasos que reducen o suprimen la capacidad del estudiante de tomar decisiones. Pocas tareas invitan al estudiante a seleccionar su propio repertorio de técnicas, recorrer una cadena de razonamientos o comparar métodos alternativos.
3. Rango de respuestas posibles. Pocas tareas ofrecen al estudiante la oportunidad de trabajar con un amplio rango de capacidades y talentos. Usualmente las tareas, más que el estudiante, determinan el rango de respuestas posible.
4. Extensión y valor de la tarea. Generalmente, el pensamiento de orden superior se manifiesta mejor en tareas largas que en tareas cortas. Es necesario que estas actividades constituyan por sí mismas experiencias de aprendizaje válidas y aceptables.
5. Modo de trabajar las tareas. Tradicionalmente, los estudiantes han trabajado las tareas individualmente y en silencio. Estas condiciones artificiales se han impuesto en beneficio de la fiabilidad y probablemente se mantendrán en el sistema. Sin embargo, hay una gran necesidad de explorar cómo se puede evaluar la capacidad de los estudiantes para trabajar cooperativamente, quizás utilizando formas de comunicación orales y prácticas en un ambiente usual de trabajo.

Entendemos que, con las evaluaciones PISA como marco de referencia, analizar los ítems utilizados por las diferentes Comunidades Autónomas en sus respectivas evaluaciones de diagnóstico permitirá identificar los esfuerzos positivos que éstas realizan para cumplir con el

marco normativo de la LOE. Por un lado, reconocer las fortalezas de los instrumentos de evaluación diseñados logrará que se mantengan y se reafirmen las prácticas positivas. De otra parte, reconocer las debilidades de esos instrumentos permitirá su reestructuración y perfeccionamiento para futuras aplicaciones.

Enmarcada en la normativa nacional, en las nociones de competencia y alfabetización matemática según están concebidas por el modelo matemático de PISA y en el enfoque funcional del aprendizaje matemático, esta investigación consta tanto de actualidad científica y educativa como de impacto social. Las evaluaciones PISA responden al marcado y creciente interés de toda una comunidad internacional en la adquisición de la competencia matemática para el desarrollo y la formación de ciudadanos reflexivos. Las evaluaciones de diagnóstico, por su parte, responden a la reciente orientación legal, por virtud de la Ley Orgánica de Educación, en el concepto de competencia y a la preocupación de toda la comunidad educativa española en la misma problemática.

Nuestro objeto de estudio es un fenómeno educativo vivo y de actualidad. Junto con el interés personal de la investigadora— profesora universitaria y consultora en programas de mejoramiento de la calidad de escuelas públicas en su país— en los procesos educativos de enseñanza, aprendizaje y evaluación, el estudio implica asimismo el interés de la comunidad educativa española en la descripción e identificación de indicadores de calidad para los instrumentos de evaluación diseñados por las administraciones educativas, interés que, a su vez, responde a las decisiones legales de una estructura gubernamental.

3. Objetivos del estudio

El objetivo general que provee orientación a esta investigación es el siguiente:

Determinar el grado de ajuste al modelo PISA de los instrumentos elaborados por las comunidades autónomas para atender a la evaluación de diagnóstico en segundo curso de la educación obligatoria.

Se pretende cumplir con el objetivo general propuesto mediante el logro de los siguientes objetivos específicos:

1. Localizar las pruebas de diagnóstico para evaluar la competencia matemática aplicadas por las comunidades autónomas a los estudiantes de 2º de la ESO en el curso 2008-2009.

2. Identificar los valores de las variables de estudio en cada uno de los ítems de acuerdo a la caracterización de las pruebas PISA.
3. Analizar los ítems de las pruebas de acuerdo a sus características y su ajuste a las evaluaciones PISA.

4. Estructura del documento

Esta investigación se presenta en cinco capítulos en el primero de los cuales se realiza una aproximación al problema bajo estudio, se justifica la investigación y se describen los objetivos que la guían. El segundo capítulo constituye el marco de referencia de la investigación. En el tercer capítulo se expone el enfoque metodológico que incluye el proceso de búsqueda y recopilación de las pruebas de diagnóstico en las comunidades autónomas, la definición de la muestra utilizada, el tipo de estudio y las técnicas de análisis así como las limitaciones del estudio. En el cuarto capítulo se presenta una descripción del análisis realizado. Finalmente, el capítulo quinto concluye la investigación con un resumen de los hallazgos principales, algunas conclusiones generales y recomendaciones para investigaciones futuras.

CAPÍTULO 2

MARCO DE REFERENCIA DE LA INVESTIGACIÓN

Este capítulo agrupa y resume los distintos fundamentos y orientaciones sobre los cuales se sostiene esta investigación: la Ley Orgánica de Educación (LOE), las evaluaciones generales y las evaluaciones autonómicas de diagnóstico, el modelo matemático de PISA, el enfoque funcional del aprendizaje matemático y la Teoría de Respuesta a los Ítems y el modelo de Rasch.

1. La Ley Orgánica de Educación (LOE)

La Ley Orgánica 2/2006, del 3 de mayo, de Educación (LOE) se fundamenta sobre tres principios. El primero de ellos, el principio de calidad y equidad, exige que todos los ciudadanos reciban una educación de calidad en todos los niveles educativos. El segundo es el principio del esfuerzo compartido. Consiste este segundo principio en lograr que todos los componentes de la comunidad educativa colaboren para alcanzar la equidad y la calidad que persigue el primer principio. Se espera que las familias, el profesorado, los centros, las administraciones educativas y la sociedad, todos como miembros de una misma comunidad educativa, desarrollen un alto sentido de compromiso que redunde en ambientes propicios para un aprendizaje de calidad y de igualdad. El tercer principio es el principio de compromiso con los objetivos de educación comunes que la Unión Europea ha formulado: el mejoramiento de la calidad y la eficacia de los sistemas educativos y de formación, mediante una mejor capacitación profesional del profesorado, el desarrollo de las aptitudes necesarias para alcanzar la meta de ser una sociedad del conocimiento, la garantía de acceso a las tecnologías de información y comunicación, un aumento en la participación de los estudios científicos, artísticos y técnicos y el aprovechamiento óptimo de los recursos disponibles aumentando la inversión en el capital humano; el acceso a los sistemas de formación mediante ambientes de aprendizaje abierto y más atractivo, la participación más activa de la ciudadanía, la igualdad de oportunidades y una mayor cohesión social; la apertura de estos sistemas al mundo exterior reforzando la cooperación europea y fomentando el espíritu emprendedor, el dominio de lenguas extranjeras, la movilidad y el intercambio (LOE, Preámbulo).

La LOE define el currículo como “el conjunto de objetivos, competencias básicas, contenidos, métodos pedagógicos y criterios de evaluación de cada una de las enseñanzas reguladas en la presente Ley” (Artículo 6). De esta manera se incorpora la noción de competencia como parte integral del currículo en todos los niveles educativos. Se considera la introducción del concepto de competencia como un cambio significativo en las normativas curriculares complementando así la orientación hacia la noción de objetivos del aprendizaje que había imperado hasta el momento (Rico y Lupiáñez, 2008). Es importante destacar que las competencias y los objetivos difieren entre sí y entrañan expectativas distintas sobre el aprendizaje de las Matemáticas (Lupiáñez, 2009).

La LOE, y los principios que la rigen, concibe la educación como un proceso de aprendizaje dinámico y permanente que se prolonga durante toda la vida. Lograr este aprendizaje vitalicio implica que el sistema educativo proveerá a los estudiantes los conocimientos y las competencias necesarias para desempeñarse eficazmente dentro de la sociedad de la cual forman parte. Sobre las ideas de aprendizaje permanente y de desarrollo de las competencias básicas la LOE introduce, como una novedad, la realización de una evaluación de diagnóstico de las competencias básicas al finalizar el segundo ciclo de la Educación Primaria y el segundo curso de la Educación Secundaria Obligatoria (ESO), respectivamente (Artículos 21 y 29). La evaluación de diagnóstico, cuya responsabilidad recae sobre las Administraciones Educativas, tendrá carácter formativo y orientador, ofrecerá información importante sobre la situación del alumnado, los centros y el sistema educativo como un conjunto y redundará en la adopción de planes de mejora de la calidad del sistema. La LOE establece que las evaluaciones de diagnóstico, en modo alguno, serán usadas para clasificar los centros. Estas evaluaciones tendrán como marco de referencia las evaluaciones generales de diagnóstico creadas igualmente por virtud de la LOE (Artículo 144). Las evaluaciones generales de diagnóstico, a cargo del Instituto de Evaluación, se ofrecerán del mismo modo, al finalizar el segundo ciclo de la Educación Primaria y al finalizar el segundo curso de la ESO. Con carácter externo y de homogeneidad, estas evaluaciones tratarán sobre las competencias básicas del currículo que los estudiantes deberán haber adquirido al finalizar la etapa correspondiente y proveerán datos representativos del estudiantado, los centros de las Comunidades Autónomas y el sistema educativo como un conjunto.

2. Las evaluaciones generales y las evaluaciones autonómicas de diagnóstico

La Ley Orgánica de Educación (LOE) 2/2006 del 3 de mayo prevé que se realicen dos procesos distintos para las evaluaciones de diagnóstico: las evaluaciones generales de diagnóstico y las evaluaciones autonómicas de diagnóstico. La LOE propone, de un lado, que el Instituto de Evaluación y las dependencias correspondientes de las Administraciones Autonómicas trabajen colaborativamente para ofrecer las evaluaciones generales de diagnóstico. Estas pruebas tienen carácter muestral y de homogeneidad, son externas a los centros y cumplen el propósito de recopilar datos representativos de los alumnos, los centros de las comunidades autónomas e igualmente, del conjunto del Estado. De otro lado, la LOE prevé que las administraciones autonómicas realicen, con carácter censal y formativo y mediante aplicación interna, las evaluaciones autonómicas de diagnóstico.

Tanto las evaluaciones generales de diagnóstico del sistema educativo como las evaluaciones de diagnóstico de las comunidades autónomas evaluarán las competencias básicas del currículo, específicamente al finalizar el segundo ciclo de la Educación Primaria y al finalizar el segundo curso de la Educación Secundaria Obligatoria (ESO). Ambos procesos evaluativos generarán compromisos de revisión y mejora educativa con base en los resultados obtenidos (Artículo 144, LOE). Mediante las evaluaciones generales de diagnóstico se obtendrán datos representativos de las comunidades autónomas y del conjunto del Estado, mientras que las evaluaciones de diagnóstico, tendrán carácter formativo y orientador para los centros y proveerán información relevante para las familias y para la comunidad educativa.

Las evaluaciones generales de diagnóstico servirán de marco de referencia para las evaluaciones de diagnóstico que cada comunidad autónoma desarrollará. No obstante, el marco de la evaluación general de diagnóstico no constituye una imposición para las evaluaciones de diagnóstico sino que cada comunidad autónoma asumirá su responsabilidad al respecto. Se espera que las evaluaciones generales de diagnóstico contribuyan a mejorar la calidad y la equidad de la educación arrojando luz sobre la situación del sistema educativo mediante la valoración de los aprendizajes de los alumnos y el impulso de procesos de innovación y mejora del sistema educativo en conjunto. Además de contribuir a mejorar la calidad educativa del sistema, las evaluaciones generales de diagnóstico tienen la finalidad de orientar políticas educativas, aumentar la eficacia y la transparencia de la educación como sistema e informar respecto del grado de adquisición y logro de las competencias básicas del currículo.

De esta manera, los resultados de las evaluaciones de diagnóstico facilitarán a las Administraciones educativas el diseño de planes de mejoramiento y la adopción de medidas específicas de apoyo educativo en el dominio de sus competencias; asimismo, los consejos escolares y el profesorado de los centros analizarán los resultados para tomar decisiones que redunden en la mejora de la educación de sus alumnos. El objetivo integral de las evaluaciones generales de diagnóstico es conocer el estado de la situación del sistema educativo respecto al aprendizaje de los estudiantes en las competencias básicas del currículo mediante la información recopilada y analizada, comprender por qué y cómo se lograron los niveles actuales y, sobre todo, provocar instancias sucesivas de reforma y mejoramiento. Las evaluaciones generales de diagnóstico tienen como objeto las competencias básicas del currículo las cuales se relacionan con contenidos que entrañan conocimientos, habilidades y actitudes que resultan adaptables y provechosos para enfrentar situaciones y problemas de la vida real.

Lo que intentan valorar las pruebas es el grado en que la escuela moldea al alumno para hacer frente a la vida y cómo lo prepara para desenvolverse como ciudadano, de una manera eficaz, en la sociedad a la cual pertenece. Es decir, que el conjunto de competencias básicas, según definidas en el currículo, se refiere a la capacidad del estudiante para aplicar los conocimientos, habilidades y actitudes adquiridos a la realidad y a la resolución de problemas enmarcados dentro del contexto de la vida cotidiana.

Las evaluaciones generales de diagnóstico estimularán en las comunidades procesos de reflexión y autoevaluación que les lleve a identificar qué aspectos específicos de su actuación son los responsables de sus logros. Es importante destacar que la información recabada mediante las evaluaciones generales de diagnóstico no será usada como instrumento punitivo, valorativo ni de rendición de cuentas de los centros.

Para cumplir los requisitos de homogeneidad, es necesario que las evaluaciones generales de diagnóstico consideren el contexto sociocultural tanto de los alumnos como de los centros a fin de explicar los resultados de las evaluaciones y las diferencias entre los centros. A tales efectos se definirá un índice socioeconómico y cultural del alumnado y de los centros educativos que permitirá la interpretación de los resultados en el marco del contexto adecuado. Junto con las variables contextuales se recogerán, mediante cuestionarios, datos sobre variables relacionadas con los recursos de los centros y de los procesos que en ellos se

llevan a cabo, que ayuden, igualmente, a interpretar los resultados. Estas variables han sido identificadas en los estudios internacionales y nacionales como las que aportan mayor poder explicativo (Instituto de Evaluación, 2009).

2.1 Condiciones técnicas de aplicación

Las muestras que serán sometidas a las evaluaciones generales de diagnóstico y que, a su vez, contestarán los cuestionarios de contexto, serán seleccionadas aleatoriamente de la población de alumnos de 4° de Educación Primaria y de 2° de Educación Secundaria Obligatoria, respectivamente. En el curso académico 2007-2008 los grupos de expertos en evaluación de las competencias básicas seleccionados para ese propósito, elaboraron una batería de ítems que fueron aplicados en la prueba piloto realizada al comienzo del curso escolar 2008-2009. Hacia finales del curso 2008-2009 se aplicaron las evaluaciones generales de diagnóstico en el nivel primario y en el curso 2009-2010, las del nivel secundario. La competencia matemática fue evaluada en ambas aplicaciones.

En lo que a estructura se refiere, estas son pruebas escritas, denominadas de papel y lápiz, en las cuales se presenta al estudiante un estímulo que describe una situación-problema en un contexto real. Asociadas al estímulo, se presentan actividades que el estudiante trabajará y que admiten diferentes formatos. Un formato frecuentemente usado y de probadas ventajas es el de prueba objetiva. Las pruebas objetivas exhiben las ventajas de que pueden incluirse en ellas un amplio rango de ítems, carecen de subjetividad y se califican con relativa rapidez y facilidad. No obstante, un señalamiento frecuente que se hace de estas pruebas es que sólo son adecuadas para evaluar aprendizajes simples tales como definiciones, datos, hechos y conceptos y se ajustan menos para evaluar niveles de pensamiento más elevado tales como la reflexión, el análisis, la crítica, el juicio valorativo y, en general, el tipo de competencia que pretenden evaluar las evaluaciones generales de diagnóstico (Instituto de Evaluación, 2009).

Por consiguiente y para subsanar las limitaciones de las pruebas objetivas, en las pruebas de la evaluación de diagnóstico se combinaron diferentes formatos de ítems, incluidos los de prueba objetiva. Cada situación-problema que constituye el estímulo de la prueba, adoptó alguno de los siguientes formatos (INECSE, 2005):

- **Pregunta cerrada.** Se presentan bajo el criterio de selección múltiple. Las preguntas pueden ser de carácter dicotómico en las cuales sólo una alternativa es correcta y las demás son erróneas o preguntas con escala de respuesta graduada en las que existe

una alternativa correcta, una o más alternativas parcialmente correctas y una o más totalmente incorrectas.

- Preguntas que exigen el desarrollo de procedimientos y la obtención de resultados. Estas son preguntas que generalmente exigen alcanzar un resultado único, aunque podrían describirse diferentes estrategias para llegar al mismo. En este caso, tanto se valora el procedimiento como el resultado, y es posible establecer diferentes niveles de corrección en la respuesta en función del grado de desarrollo competencial evidenciado.
- Preguntas abiertas que admiten respuestas diversas. Las respuestas a estas preguntas, aunque correctas, pueden diferir de un alumno a otro. Este formato pudiera carecer de objetividad por parte del corrector. Para reducir la introducción de subjetividad en el proceso de corrección y calificación, las preguntas serán valoradas por más de un juez e irán acompañadas de una pauta de corrección que incluirá todas las posibles respuestas, estableciendo cuáles se considerarán correctas y cuáles no. La elaboración de criterios de corrección cumple el objetivo de graduar las respuestas estableciendo niveles de ejecución intermedios entre la respuesta correcta y la incorrecta.

3. El modelo matemático de PISA

El *Programme for International Student Assessment* (PISA) se considera el programa internacional más completo y riguroso para evaluar el desempeño de los estudiantes y para recopilar información respecto de los factores relacionados con el estudiantado, las familias y las instituciones que permiten explicar las diferencias en tal desempeño (OCDE, 2004). El dominio matemático que evalúan las pruebas PISA se conoce como alfabetización matemática (*mathematical literacy*) y, en términos generales, competencia matemática. Se describe este dominio como “la capacidad del estudiante para analizar, razonar y comunicar de una manera eficaz cuando enuncian, formulan y resuelven problemas matemáticos en una variedad de dominios y situaciones” (Rico, 2003, p. 276). Si el estudiante se desempeña adecuadamente en estas capacidades se dice que está *matemáticamente alfabetizado* (Rico, 2006). Esta definición gira en torno a usos más amplios de las matemáticas en la vida de las personas y no está limitada a las operaciones mecánicas o algorítmicas.

El término alfabetización matemática se usa para indicar la habilidad del alumno para usar el conocimiento y las destrezas matemáticas más que para dominarlas dentro del currículo

escolar. Quiere decir que sólo conocer las reglas, los procedimientos y los algoritmos no hace a un estudiante matemáticamente alfabetizado. Involucrarse con las matemáticas incluye no sólo realizar cálculos simples sino usos más amplios tales como asumir un punto de vista y apreciar cuestiones expresadas numéricamente, como por ejemplo emitir una opinión respecto del plan de gastos gubernamental.

Un alumno matemáticamente letrado o alfabetizado comprende y es consciente del papel que desempeñan las matemáticas en el mundo moderno. Para alcanzar este estado de plena conciencia el alumno debe ser expuesto a aplicaciones matemáticas relacionadas con la vida diaria que le permitan desarrollar la habilidad y la confianza para razonar, analizar e interpretar situaciones cotidianas y para resolver problemas enmarcados en contextos conocidos. El presentar problemas en contextos reales logra que el estudiante realice conexiones e integre todo un cúmulo de conocimientos. Por ejemplo, un estudiante puede abordar un problema algebraicamente, mientras otro estudiante, decidirá abordarlo geoméricamente.

Contrario al enfoque tradicional de las matemáticas escolares, la alfabetización matemática tiene un carácter menos formal y más intuitivo, menos simbólico y más concreto. Además, su foco de atención son, entre otras, las competencias de razonar, pensar e interpretar. Puesto que no son exclusivas de ningún área curricular específica, la competencia matemática facilita el proceso de seleccionar las estrategias adecuadas para la resolución de problemas en distintos contextos (De Lange, 2006). Por lo tanto, las situaciones reales que se presenten a los estudiantes deben ser relevantes para sus intereses.

Inherente al concepto de competencia se halla la necesidad de que el estudiante sea capaz de adaptar y aplicar su comprensión matemática a una amplia gama de contextos dentro y fuera de la sala de clase. Lejos de enfatizar el dominio de conceptos y técnicas, la evaluación de PISA enfatiza la capacidad del estudiante para enfrentarse, mediante el uso de las matemáticas, a una amplia variedad de situaciones cotidianas. Es decir, su desempeño como ciudadanos capaces de interactuar con el mundo que les rodea y de resolver matemáticamente las distintas situaciones que a diario les presenta el mundo en que viven. Por lo tanto, el foco de atención de PISA no está en lo que el estudiante ha aprendido mediante un currículo escolar específico y convencional sino en las destrezas y los conocimientos necesarios para actuar como un adulto informado, reflexivo y como miembro de una sociedad. La

alfabetización matemática se evalúa aplicando a los estudiantes tareas basadas en situaciones que representan el tipo de problemas que enfrentarán en la vida real.

Dentro del marco conceptual del modelo matemático PISA se denomina matemización a la estrategia general que usan los estudiantes para resolver problemas de la vida real. Matematizar es un proceso fundamental e inherente a la alfabetización matemática (OCDE, 2009). Cuando un estudiante se mueve entre el contexto real del problema y el mundo matemático necesario para resolverlo, está matematizando. Este proceso cíclico incluye tanto la interpretación y la evaluación del problema como la reflexión en su solución para asegurarse que la solución obtenida, en efecto, responde a la situación real que inicialmente lo originó. La actividad de matematización puede caracterizarse mediante cinco componentes (OCDE, 2009):

1. Presentar un problema enmarcado en la vida real
2. Organizar el problema en función de unos conceptos matemáticos e identificar la información matemática relevante.
3. Descartar la información irrelevante para centrarse en los aspectos matemáticos de la información mediante procesos tales como formular presunciones, generalizar y formalizar.
4. Resolver el problema.
5. Reflexionar y dar sentido a la solución matemática en términos de la situación real incluyendo identificar las limitaciones de la solución.

Estos cinco componentes pueden agruparse en tres fases de acuerdo a los aspectos generales de enfoques conocidos en el proceso de resolución de problemas (por ejemplo, Polya, 1962).

Fase 1. Comprender el problema. Este proceso también se conoce como matematización horizontal.

Fase 2. Refinar la aproximación al problema. Este proceso también se conoce como matematización vertical.

Fase 3. Interpretar la solución matemática. Incluye relacionar las respuestas al contexto.

El siguiente diagrama ilustra el ciclo del proceso de matematización.

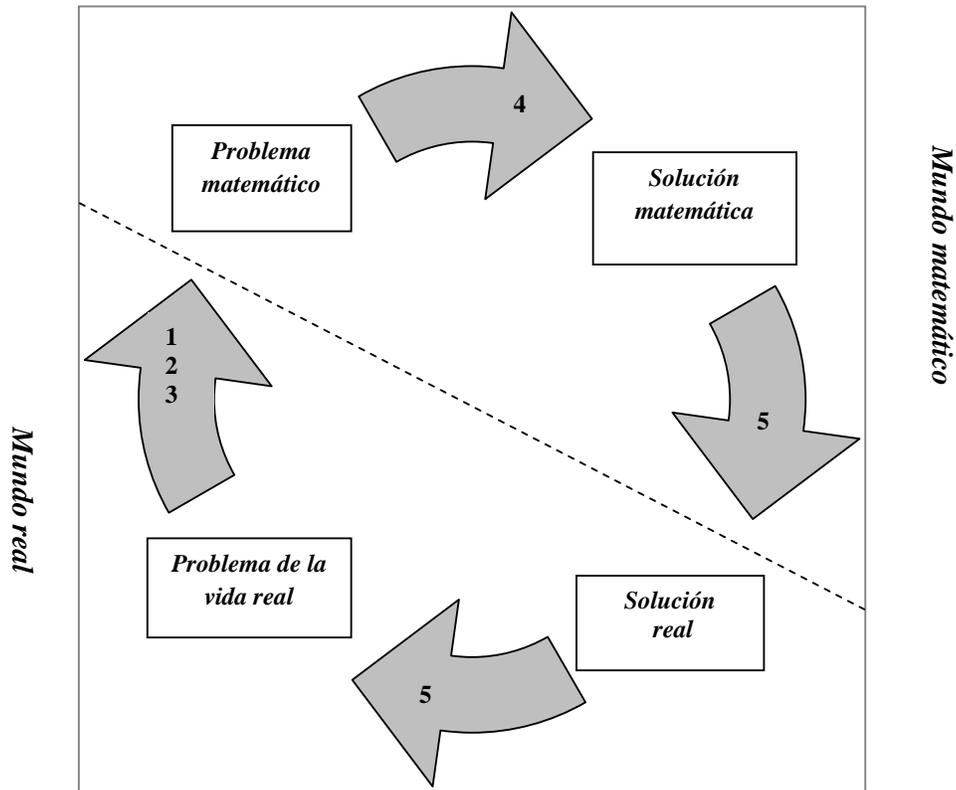


Figura 2. El ciclo de la matematización

3.1. Caracterización de las pruebas PISA

Las pruebas PISA se ofrecen a estudiantes que hayan cumplido 15 años sin tomar en consideración el nivel escolar que cursen en ese momento. Estos estudiantes se encuentran cerca del final de la educación obligatoria y en el umbral de la vida adulta, cuando se enfrentan a la decisión de incorporarse a la fuerza trabajadora o de continuar estudios más avanzados. La razón principal para considerar una edad biológica y no un grado académico descansa en la dificultad de encontrar suficientes puntos comunes entre los países participantes como para establecer comparaciones de los resultados (De Lange, 2006). El optar por una edad biológica en lugar de un determinado curso académico ha orientado las pruebas hacia un enfoque no academicista centrandose su atención, no en comprobar si el estudiante ha aprendido lo que se le ha enseñado sino en si es capaz de activar los conocimientos y destrezas adquiridas para enfrentarse con éxito a la vida adulta. En la caracterización de las evaluaciones matemáticas PISA se distinguen variables de tarea, variables de sujeto y variables de resultado de desempeño (OCDE, 2004).

3.1.1. Variables de tarea

Para satisfacer el principio de matematización o competencia sobre el cual se fundamentan, los instrumentos de evaluación de las pruebas PISA consideran tres dimensiones o variables de tarea: el contenido matemático al que se refieren las tareas, las situaciones o contextos usados como estímulo para enmarcar el problema y la complejidad de las competencias que se esperan activar para relacionar la tarea con el mundo real, agrupadas éstas en niveles de demanda cognitiva.

1. Contenidos matemáticos. Atemporadas a la orientación principal del currículo escolar, PISA considera las siguientes ideas principales (*overarching ideas*) como contenidos matemáticos consideradas categorías fenomenológicas:
 - a. Cantidad. Este contenido responde a la Aritmética y abarca todo lo relativo a la cuantificación necesaria para organizar el mundo. En esta categoría se incluye el sentido numérico, la comprensión del significado de las operaciones, el sentido de la magnitud de los números, los cálculos elegantes, los cálculos mentales y las estimaciones.
 - b. Espacio y forma. Este contenido responde a la Geometría tradicional e implica comprender, describir e interpretar el mundo físico que nos rodea. Se incluyen en esta categoría los siguientes aspectos: reconocer formas y modelos, describir, codificar y decodificar la información visual, comprender los cambios dinámicos de las formas, establecer similitudes y diferencias, posiciones relativas, representaciones bidimensionales y tridimensionales y las relaciones entre ambas y la orientación en el espacio.
 - c. Cambio y relaciones. Este contenido responde al Álgebra y cubre las manifestaciones y los procesos de cambio en el mundo que nos rodea abordados mediante el estudio de las relaciones que, a su vez, se abordan desde el ámbito de las funciones matemáticas. Para representar las relaciones pueden usarse los símbolos, las tablas y las gráficas y los dibujos geométricos.
 - d. Incertidumbre. Este contenido responde a la Estadística y la Probabilidad y abarca el tratamiento de los datos y el azar. Las ideas principales que se

incluyen son la recopilación, análisis y representación de los datos, la probabilidad y la inferencia.

2. Contextos. Considerado un aspecto importante y relevante en el proceso de matematización, el contexto permite al estudiante ubicar la tarea propuesta en una situación que le resulta familiar. Según su grado de cercanía con la situación particular del estudiante, se distinguen cuatro tipos de contextos:
 - a. Contextos personales. Se relacionan con las actividades cotidianas que tienen relevancia personal directa e inmediata para el estudiante.
 - b. Contextos educativos o laborales. Son situaciones a las cuales el estudiante pudiera enfrentarse en el ambiente escolar o en un entorno de trabajo. Incluye tareas que propone el profesor con fines exclusivamente para la instrucción.
 - c. Contextos públicos. Son situaciones que surgen en la interacción diaria del individuo con el mundo externo.
 - d. Contextos científicos. Son situaciones más abstractas con las cuales el estudiante está poco relacionado. Puede incluir un proceso tecnológico, una teoría o un problema eminentemente matemático.
3. Niveles de complejidad. Cada una de las competencias que se trabajan en la evaluación PISA plantea distintos niveles de demanda cognitiva o diferentes niveles de profundidad. En el marco PISA (OCDE, 2004), esta variable se denomina “conjuntos de competencia (*competency clusters*). En este trabajo adoptamos la caracterización realizada por Rico y Lupiáñez (2008) que concretan esta variable en términos de niveles de complejidad de las tareas. En general, estas competencias están asociadas con tareas cuya dificultad va en ascenso; no obstante, existe superposición en la clasificación de las tareas dentro de cada categoría. Así, se distinguen los siguientes:
 - a. Reproducción. Las preguntas en este nivel requieren que el estudiante demuestre que domina el conocimiento aprendido. Son problemas que les resultan familiares y se resuelven aplicando algoritmos o destrezas técnicas. Incluye los procesos de acceder (recordar, reproducir) e identificar.

- b. **Conexión.** Las preguntas en este nivel requieren que el estudiante demuestre que puede establecer relaciones entre distintos dominios matemáticos y que puede integrar información para resolver problemas que no son rutinarios pero que exigen que el estudiante se decida por una de entre varias estrategias de resolución. Incluye los procesos de aplicar, analizar y valorar.
- c. **Reflexión.** Las preguntas en este nivel son situaciones poco estructuradas que requieren que el estudiante comprenda, reflexione y use su creatividad para reconocer las matemáticas involucradas en el problema. Se exige que el estudiante analice, interprete y desarrolle sus propios modelos y estrategias y presente argumentos matemáticos, demostraciones y generalizaciones. Incluye los procesos de sintetizar, crear y juzgar.

3.1.2. Variables de sujeto

Las variables de sujeto consideradas en PISA son aquellas variables independientes que caracterizan a la muestra elegida: edad, sexo, nacionalidad, Comunidad Autónoma, enseñanza pública o privada, nivel socioeconómico, nivel cultural, etc. Estas variables son las propias de la población española. Algunas variables, como la edad, adoptan valores fijos mientras que otras sí toman valores distintos. Esta información suele recogerse en las tablas que describen la muestra en la aplicación de cada una de las pruebas. Para este estudio estas variables no tienen significado especial ya que nuestros sujetos son los ítems de las pruebas y no los “sujetos” en el sentido de alumnos.

3.1.3. Variables de resultado o de desempeño

El uso efectivo del conocimiento, destrezas, habilidades y actitudes matemáticas por un individuo al abordar cuestiones y problemas en una variedad de contextos muestra su competencia matemática. PISA es un dominio de gran complejidad. Para llevar a cabo un análisis detallado de la competencia matemática se han identificado ocho componentes o competencias características específicas que son relevantes y significativas en todos los niveles escolares. Enfatizamos que estas competencias no responden a áreas o contenidos curriculares específicos. Puede apreciarse que estas competencias se superponen unas a las otras por el grado de interacción que se manifiesta entre ellas al aplicar el conocimiento matemático en la resolución de problemas. Quiere decir que un alumno puede mostrar

diferentes niveles de dominio sobre una o varias de estas competencias, simultáneamente, al enfrentarse a una situación-problema.

Estas competencias y niveles proveen la base para las escalas o niveles de rendimiento que constituyen la variable de resultado o desempeño en el estudio PISA. Con el propósito de determinar el desempeño específico del estudiante en las pruebas, se han identificado seis niveles de competencia o rendimiento (IE, 2006). Estos niveles permiten conocer el grado de competencia matemática que poseen los alumnos.

1. Competencias matemáticas específicas. Más que a conceptos o destrezas curriculares, el término se refiere a lo que el estudiante es capaz de hacer con los conceptos y destrezas aprendidos. El estudio PISA distingue las siguientes (OCDE, 2009):
 - a. Pensar y razonar. Considerada una competencia matemática fundamental, la capacidad para pensar y razonar incluye formular preguntas exploratorias respecto de lo que es posible o lo que puede ocurrir bajo determinadas circunstancias, cómo proceder para investigar alguna situación y analizar lógicamente las conexiones entre los elementos de un problema.
 - b. Argumentar. Incluye la habilidad para seguir cadenas de razonamientos y argumentos así como crear tales cadenas para analizar matemáticamente un proceso. Esta competencia también puede surgir para explicar, justificar o verificar un resultado y se considera medular para la alfabetización matemática.
 - c. Comunicar. Incluye la comprensión de comunicaciones escritas, orales y gráficas de segundas personas así como la capacidad para expresar el propio punto de vista matemático sobre diversos asuntos.
 - d. Modelar. Considerada una competencia crítica para la alfabetización matemática, la competencia de modelar se refiere a la capacidad del individuo para moverse cómodamente entre el mundo real y el mundo matemático.
 - e. Plantear y resolver problemas. Definir y clarificar el problema es un paso importante para su resolución. Esta competencia incluye plantear, formular y

definir diferentes tipos de problemas matemáticos y el uso de estrategias variadas para su solución.

- f. Representar. Es una competencia básica pero importante para la alfabetización matemática. Incluye la habilidad para usar y manipular de forma exitosa distintas representaciones de objetos y situaciones matemáticas tales como gráficas, tablas, fotografías, diagramas, textos, representaciones algebraicas y otras representaciones matemáticas.
 - g. Utilizar el lenguaje simbólico, formal y técnico y las operaciones. Incluye decodificar el lenguaje simbólico y comprender su relación con el lenguaje natural. Esta competencia también se relaciona con la habilidad para manejar y trabajar con enunciados que contengan símbolos y fórmulas. Incluye además el dominio de destrezas técnicas y de procedimiento asociadas a una variedad de procesos matemáticos formales.
 - h. Usar ayudas y herramientas. Esta competencia está asociada con el conocimiento y la capacidad del estudiante para usar diferentes ayudas y herramientas tales como herramientas de información tecnológica. El estudiante necesita reconocer cuándo estas herramientas pueden ser de utilidad y qué limitaciones poseen así como saber usarlas de manera apropiada.
2. Niveles de rendimiento (Instituto de Evaluación, 2006, pp. 4-5)
- a. Nivel 6. En este nivel *los alumnos saben formar conceptos, generalizar y utilizar información basada en investigaciones y modelos de situaciones de problemas complejos. Pueden relacionar diferentes fuentes de información y representaciones y traducirlas de una manera flexible. Los estudiantes de este nivel poseen un pensamiento y razonamiento matemático avanzado. Estos alumnos pueden aplicar su entendimiento y comprensión, así como su dominio de las operaciones y relaciones matemáticas simbólicas y formales y desarrollar nuevos enfoques y estrategias para abordar situaciones nuevas. Los alumnos pertenecientes a este nivel pueden formular y comunicar con exactitud sus acciones y reflexiones relativas a sus descubrimientos, argumentos y su adecuación a las situaciones originales.*

- b. Nivel 5. En este nivel *los alumnos saben desarrollar modelos y trabajar con ellos en situaciones complejas, identificando los condicionantes y especificando los supuestos. Pueden seleccionar, comparar y evaluar estrategias adecuadas de solución de problemas para abordar problemas complejos relativos a estos modelos. Los alumnos pertenecientes a este nivel pueden trabajar estratégicamente utilizando habilidades de pensamiento y razonamiento bien desarrolladas, así como representaciones adecuadamente relacionadas, caracterizaciones simbólicas y formales, e intuiciones relativas a estas situaciones. Pueden reflexionar sobre sus acciones y formular y comunicar sus interpretaciones y razonamientos.*
- c. Nivel 4. En este nivel *los alumnos pueden trabajar con eficacia con modelos explícitos en situaciones complejas y concretas que pueden conllevar condicionantes o exigir la formulación de supuestos. Pueden seleccionar e integrar diferentes representaciones, incluyendo las simbólicas, asociándolas directamente a situaciones del mundo real. Los alumnos de este nivel saben utilizar habilidades bien desarrolladas y razonar con flexibilidad y con cierta perspicacia en estos contextos. Pueden elaborar y comunicar explicaciones y argumentos basados en sus interpretaciones, argumentos y acciones.*
- d. Nivel 3. En este nivel *los alumnos saben ejecutar procedimientos descritos con claridad, incluyendo aquellos que requieren decisiones secuenciales. Pueden seleccionar y aplicar estrategias de solución de problemas sencillos. Los alumnos de este nivel saben interpretar y utilizar representaciones basadas en diferentes fuentes de información y razonar directamente a partir de ellas. Son también capaces de elaborar breves escritos exponiendo sus interpretaciones, resultados y razonamientos.*
- e. Nivel 2. En este nivel *los alumnos saben interpretar y reconocer situaciones en contextos que sólo requieren una inferencia directa. Saben extraer información pertinente de una sola fuente y hacer uso de un único modelo de representación. Los alumnos de este nivel pueden utilizar algoritmos,*

fórmulas, procedimientos o convenciones elementales. Son capaces de efectuar razonamientos directos e interpretaciones literales de los resultados.

- f. Nivel 1. En este nivel los alumnos saben responder a preguntas relacionadas con contextos que les son conocidos, en los que está presente toda la información pertinente y las preguntas están claramente definidas. Son capaces de identificar la información y llevar a cabo procedimientos rutinarios siguiendo unas instrucciones directas en situaciones explícitas. Pueden realizar acciones obvias que se deducen inmediatamente de los estímulos presentados.

Una vez normalizadas, las puntuaciones alcanzadas por cada uno de los estudiantes en las pruebas PISA identifican cada uno de estos niveles mediante un intervalo de calificaciones. El desempeño de un estudiante puede señalarse por una valoración numérica y describirse por su nivel de desarrollo en las competencias matemáticas específicas, cuyos descriptores se acaban de presentar.

A los efectos de nuestro estudio, centrado en los instrumentos de evaluación de las pruebas de diagnóstico, son las variables de tarea que delimitan el diseño de los ítems las que tienen mayor relieve. Las variables de sujeto deben atender a respetar las condiciones y requisitos generales del estudio PISA y a satisfacer los criterios de representatividad de las muestras. Las variables de resultado permitirán, en su caso, efectuar un control empírico de idoneidad de los instrumentos de evaluación.

4. Las variables de tarea en PISA como criterio para evaluar los ítems

La estructura del modelo matemático PISA puede caracterizarse mediante la siguiente representación matemática: AM + 3Cs. En esta expresión AM significa *alfabetización matemática* y las tres Cs representan las variables de tarea: contexto, contenido y complejidad. Los problemas propuestos ocurren en una situación real; esta situación provee el contexto para la tarea matemática. A fin de usar las matemáticas para resolver el problema, el estudiante debe poseer un grado de dominio del contenido matemático relevante. Y a fin de resolver el problema debe desarrollarse y seguirse un proceso de resolución del problema. Luego, para ejecutar este proceso exitosamente, el estudiante necesita determinadas competencias que el modelo agrupa al establecer tres niveles hipotéticos de complejidad en la

demanda cognitiva que plantea la tarea. De esta manera el estudiante matemáticamente alfabetizado se enfrenta a un proceso de resolución de un problema enmarcado en un contexto, un contenido matemático y mediante una demanda que tiene un determinado nivel de complejidad. Vemos cómo el modelo PISA se ajusta al modelo funcional del conocimiento matemático conformado por unas tareas contextualizadas, unas herramientas conceptuales y el sujeto cognitivo que trabaja (OCDE, 2009).

En PISA el contenido no se considera como relativo a las áreas matemáticas tradicionales sino como ideas amplias o clave (*overarching ideas*) estructuradas alrededor de diferentes fenómenos que describen conceptos, ideas y estructuras matemáticas. Es decir, son categorías de carácter fenomenológico, relacionadas con áreas más tradicionales de las matemáticas, que describen el contenido matemático en relación con los fenómenos y la variedad de problemas para los que fueron creados. No existe una correspondencia estricta entre estas ideas amplias y las áreas matemáticas tradicionales. Lo que PISA intenta valorar es si la capacidad de los estudiantes para profundizar es suficiente como para identificar los conceptos matemáticos y demostrar una mejor comprensión del significado de estos conceptos en el mundo.

Frecuentemente subestimado o ignorado en las matemáticas escolares, el contexto en el cual se sitúa un problema juega un papel importante en la resolución de problemas de la vida real y en la alfabetización matemática. PISA destaca la importancia del contexto y le otorga un papel principal en la evaluación de la alfabetización matemática. Puesto que el ciudadano común se relaciona con contextos de diversa naturaleza, PISA reconoce la necesidad de incluir en sus evaluaciones una amplia gama de contextos. El contexto no ha sido ampliamente estudiado y no existen resultados concluyentes respecto de cuáles contextos resultan más atractivos para los estudiantes o más apropiados para las evaluaciones (OCDE, 2009). No obstante, una creencia generalizada es que, particularmente en los grados primarios, los estudiantes prefieren los contextos más cercanos al ambiente en el que se mueven porque los asimilan más fácilmente. Esto no es necesariamente cierto para los estudiantes en el nivel secundario puesto que su comprensión del mundo es más amplia y ya se relacionan con contextos de otra índole como los científicos y los políticos.

En términos de su relevancia matemática, PISA distingue entre contexto camuflado o de orden cero, contexto de primer orden y contexto de segundo orden. El contexto es camuflado si sólo se usa para que el problema matemático parezca enmarcado en el mundo real. En estos casos, el contexto se considera sólo un pretexto y nada de lo que concierne a éste, es

necesario para comprender o resolver el problema. PISA prefiere mantener sus evaluaciones alejadas de este tipo de contexto aunque han sido usados en contadas ocasiones. Se considera de primer orden al contexto que es relevante y necesario para resolver el problema y para juzgar la razonabilidad de su respuesta. Si el estudiante debe moverse entre la situación del contexto y el problema matemático para resolverlo o si debe enmarcarse en el contexto para reflexionar en la solución, entonces hay un proceso claro de matematización y el contexto se considera de segundo orden. Los niveles más altos de alfabetización matemática requieren el manejo de contextos de segundo orden (OCDE, 2009).

El ciudadano que usa de manera efectiva el conocimiento matemático adquirido en una variedad de contextos necesita dominar ciertas competencias matemáticas. Cuando se enfrenta a una tarea, el estudiante moviliza distintas capacidades y a diferentes niveles de ejecución. Las ocho competencias definidas en PISA– pensar y razonar, argumentar, comunicar, modelizar, representar, plantear y resolver problemas, utilizar el lenguaje formal, simbólico y técnico y las operaciones y usar ayudas y herramientas– son agrupadas en tres niveles cognitivos de complejidad: reproducción, conexión y reflexión que ya discutimos anteriormente.

5. El enfoque funcional del aprendizaje matemático

La evaluación PISA concibe las matemáticas como un conjunto de herramientas que sirven para resolver problemas mediante la puesta en funcionamiento de determinadas competencias (Rico, 2006). Tanto las nociones de alfabetización y competencia según se describe en el modelo PISA, como la percepción de las matemáticas como “modo de hacer” se ajustan a un enfoque funcional del aprendizaje de las matemáticas. Rico y Lupiáñez (2008) establecen que:

En el enfoque funcional de las matemáticas escolares los conceptos y procedimientos matemáticos tienen un para qué cercano, sirven para algo tangible, pues las nociones matemáticas constituyen herramientas mediante las que actuamos para dar respuesta a cuestiones, problemas e interrogantes del entorno (p. 178).

Este modelo funcional del aprendizaje matemático no centra su foco de atención en los conocimientos adquiridos mediante los contenidos curriculares sino en cómo los alumnos

aplican tales conocimientos para enfrentarse a situaciones de la vida real que le son familiares. Son tres las dimensiones que conforman el modelo funcional: tareas contextualizadas, herramientas conceptuales y el sujeto cognitivo que trabaja. ¿Cómo interactúan estas tres dimensiones? El sujeto cognitivo usa las herramientas que tiene a su disposición para aproximarse a las tareas, movilizándolo y manifestando su competencia al efectuar los correspondientes procesos cognitivos.

El modelo funcional requiere que el estudiante conozca las herramientas que tiene disponibles, pero su foco de atención principal reside en las tareas. Este primer acercamiento en el dominio de cómo es la herramienta responde a un enfoque instrumental del aprendizaje matemático. Contrario al modelo funcional, el modelo instrumental centra su atención en los contenidos curriculares básicos de las matemáticas escolares. No obstante, puede afirmarse que ambos enfoques coexisten, son complementarios o consecutivos. La analogía que se establece es la siguiente: una vez el sujeto conoce las herramientas básicas de las matemáticas (conceptos y objetos), el énfasis se dirige, no a la herramienta misma como objeto de estudio, sino a su aplicación a distintas situaciones, a su empleo en satisfacer las demandas que plantean las tareas. Es decir, en el uso de la herramienta en situaciones no convencionales. La orientación de funcionalidad dentro del currículo de Matemáticas no implica el conocimiento de los conceptos básicos, como herramientas, en todas sus manifestaciones sino la extensión de su uso y aplicación a situaciones que así lo demanden (Rico y Lupiáñez, 2008).

6. La Teoría de Respuesta a los Ítems y el modelo de Rasch

La competencia que poseen los estudiantes que responden a una prueba específica se estima tomando en consideración la proporción de ítems que contestan correctamente de acuerdo a la dificultad de los ítems. Por su parte, la dificultad relativa de los ítems de una prueba se estima de acuerdo a la proporción de estudiantes que aciertan cada ítem. El análisis de los datos que se recopilan en las pruebas PISA se realiza mediante un modelo matemático denominado la Teoría de Respuesta al Ítem (TRI) (Lord, 1990).

Más poderosa que y contraria a la Teoría Clásica de los Tests (TCT) postulada en los años 20, la TRI se enfoca más en las características de los ítems de una prueba que en la prueba como un todo o en las puntuaciones obtenidas (Hambleton & Hariharan, 1985). Esta teoría considera al ítem como la unidad básica de medición. Mediante la TCT, la medición estaba sujeta al instrumento de medida utilizado y los sujetos evaluados definían las características

de tales instrumentos. Los tests elaborados usando la TCT como marco de referencia, tenían la limitación de que las características del examinando no podían separarse de las características del test (Cortada de Kohan, 2004). En cambio, en la TRI las mediciones son independientes de los instrumentos y las propiedades de estos son invariantes respecto de los sujetos evaluados. Es decir, que las propiedades de la prueba como un instrumento de medida, se expresan en términos de las propiedades de los ítems de la prueba.

Para medir una habilidad específica, se desarrolla una prueba que consiste de ítems o preguntas cada uno de los cuales mide un aspecto de la habilidad en cuestión. El ítem es una variable dicotómica: se otorga 1 a la respuesta correcta y 0 a cada uno de los distractores. Se presume que cada examinando posee algún grado de la habilidad que se examina en la prueba. Por lo tanto, se considera que cada examinando tiene una puntuación o valor numérico que lo coloca en algún punto de la escala de habilidad. Esta habilidad se denota por la letra griega θ . Se denota por $P(\theta)$ a la probabilidad de que un examinando con un determinado nivel de habilidad responda a un ítem correctamente. La probabilidad de una respuesta correcta está muy cerca de 0 en los niveles más bajos de habilidad y aumenta hasta que en los niveles más altos de habilidad, la probabilidad de una respuesta correcta se acerca a 1. Si se grafica $P(\theta)$ como una función de la habilidad, se obtiene una curva suave en forma de S. En la TRI esta curva se conoce como la curva característica del ítem. Cada ítem en una prueba tiene su propia curva característica (Baker, 2001).

La TRI se basa sobre los siguientes dos supuestos (Muñiz & Fidalgo, 2005):

1. La unidimensionalidad del rasgo latente. Quiere decir que los ítems que conforman un test deben medir sólo una aptitud o destreza.
2. La independencia. Este supuesto significa que las respuestas a cualquier par de ítems son independientes y no existe relación entre las respuestas a diferentes ítems.

Dentro de la TRI se distinguen distintos modelos de acuerdo al número de parámetros que utilizan para describir los ítems. El modelo que se aplica a las evaluaciones de PISA es el modelo logístico de un sólo parámetro conocido como el modelo de Rasch. Se designa este modelo como de parámetro único porque las curvas características de los ítems sólo dependen de la dificultad de los mismos. Investigadores como Curtis y Boman (2007) aseguran que usar el modelo de Rasch en el desarrollo de instrumentos de evaluación provee al investigador información de diagnóstico más útil y detallada.

Puesto que en nuestro estudio no trabajamos con los resultados individuales de las pruebas sino con la redacción de los ítems y su aplicación a estas, el uso del modelo de Rasch queda pospuesto para ser aplicado a un posible trabajo empírico en el futuro.

CAPÍTULO 3

METODOLOGÍA DE INVESTIGACIÓN

Como se explica en el capítulo 1, este estudio trata sobre las pruebas elaboradas para la primera evaluación de diagnóstico en competencia matemática realizada en España por las Consejerías de Educación de las Comunidades Autónomas a los estudiantes de segundo curso de la Educación Secundaria Obligatoria (ESO) en el año académico 2008-2009. Este capítulo presenta el enfoque metodológico de la investigación. Se describe el proceso de búsqueda y recopilación de las pruebas, la definición de la muestra, el tipo de estudio, las técnicas de análisis y las limitaciones del estudio.

1. Búsqueda y recopilación de las pruebas de diagnóstico

El acceso a las pruebas censales de evaluación se logró mediante una búsqueda sistemática en los portales de las Consejerías de Educación de cada comunidad. Como resultado de esa búsqueda se logró acceso a cinco de las pruebas aplicadas. Del total de 17 comunidades autónomas, sólo cuatro tienen disponible la prueba en su página web por lo que se procedió a establecer comunicación con los encargados de las evaluaciones de diagnóstico en las comunidades restantes. Como resultado de tal gestión, se logró acceso a una prueba adicional. Hubo dos comunidades que, pese a varios intentos, no respondieron a la solicitud; tres comunidades sólo permitían acceso a la prueba aplicada en 4º de primaria bien porque no aplicaron prueba de competencia matemática en 2º de la ESO o porque no estaba disponible para divulgación; tres comunidades indicaron que la prueba no estaba disponible para divulgación porque sería usada para futuras aplicaciones, dos comunidades indicaron no haber evaluado competencia matemática en 2º de la ESO, una comunidad publicó sólo una parte de la prueba como ítems liberados; las pruebas de una comunidad serían liberadas en una fecha posterior y una comunidad sólo publicó ejemplos de ítems. La figura 3 resume la disponibilidad de las pruebas.



Figura 3. Disponibilidad de las pruebas por Comunidad Autónoma

En resumen, nuestro banco de información consta de cinco pruebas. Para garantizar la confidencialidad de las comunidades autónomas se asignó aleatoriamente un código a cada una de ellas. En lo sucesivo nos referiremos a ellas como comunidad A, comunidad B, comunidad C, comunidad D y comunidad E. Es importante destacar que la clasificación final de las comunidades no coincide con su organización alfabética.

2. Definición de la muestra

La muestra de ítems incluida en este trabajo es una muestra por disponibilidad. A pesar de los múltiples intentos realizados para lograr acceso a más pruebas, finalmente limitamos la muestra a los ítems procedentes de las cinco pruebas descritas anteriormente. Puesto que cinco comunidades no ofrecieron pruebas de evaluación de diagnóstico en el curso 2008-2009, el banco de datos obtenido pudiera calificarse como suficiente, ya que comprende un porcentaje estimado del 50% de las comunidades que aplicaron pruebas. No obstante, nuestro objeto de estudio son los ítems individuales; por lo tanto, para determinar la fortaleza de la muestra en términos del porcentaje de ítems sobre el total, necesitábamos conocer el total de ítems incluido en las pruebas no disponibles, información que solicitamos a las Consejerías de Educación de las comunidades. Excepto la Comunidad Autónoma de Canarias, las dependencias correspondientes de las comunidades a cuyas pruebas no tuvimos acceso no

atendieron nuestra solicitud, por lo cual desconocemos qué porcentaje del total de ítems aplicados representa nuestra muestra.

En resumen, las pruebas disponibles contienen un total de 173 ítems distribuidos como se muestra en la Tabla 1.

Tabla 1
Distribución del total de ítems por Comunidad Autónoma

Comunidad Autónoma	Número de ítems	Porcentaje
A	18	10,4
B	84	48,6
C	32	18,5
D	25	14,4
E	14	8,1
Total	173	100,0

Puede observarse que no existe pauta alguna en lo que a número de ítems se refiere. La comunidad B incluyó la mayor cantidad de ítems con 84, representando esta cantidad cerca de la mitad (48,6 %) del total de ítems en la muestra. Por su parte, la comunidad E contenía sólo 14 ítems, el menor número de ellos (8,1 %).

3. Tipo de estudio

Calificamos nuestro estudio como exploratorio-descriptivo. Esta es la primera investigación de la cual tenemos conocimiento que atiende la aplicación inicial de las pruebas autonómicas de diagnóstico. Aunque enmarcada en el modelo matemático de las evaluaciones de PISA, esta investigación carece de estudios previos que permitan establecer comparaciones o constatar los hallazgos. Este hecho le otorga a nuestro estudio el carácter de exploratorio.

De otro lado, nuestro propósito es documentar el objeto de estudio, los ítems individuales de las pruebas, de la manera más completa y exhaustiva posible y, como propósito de toda investigación, servir de marco de referencia para otras investigaciones. Para lograrlo caracterizamos los ítems de las pruebas disponibles de acuerdo a las variables de tarea del modelo de PISA y al modelo funcional del aprendizaje matemático. Queremos ofrecer una

visión general de los ítems diseñados y aplicados por las comunidades autónomas mediante la identificación de sus características de acuerdo a las variables contenido, contexto y nivel de complejidad. Este hecho le otorga a nuestro estudio el carácter de descriptivo.

4. Técnicas de análisis

Por la naturaleza exploratorio-descriptiva de la investigación, el análisis de los hallazgos se realiza mediante técnicas descriptivas. Luego de caracterizar los ítems de las pruebas de acuerdo a las variables contenido, contexto y nivel de complejidad, se calculan porcentajes, se organizan los hallazgos en tablas de frecuencias y se representan mediante gráficos. Con el propósito de observar la relación entre las variables, se construyen tablas de contingencia tanto para las tres variables de estudio como para parejas de variables. La relación que existe entre ellas y que se observa en una tabla de contingencia, se comprueba mediante un análisis logarítmico lineal. El análisis log lineal es una prueba estadística de dependencia. Esta prueba se aplica en situaciones, como es nuestro caso, donde existen dos o más variables categóricas o nominales y se necesita saber si existen relaciones de dependencia entre ellas.

5. Limitaciones del estudio

Reconocemos que la investigación presenta dos limitaciones. La primera limitación concierne a la disponibilidad de las pruebas. No es aceptable que se haya negado el acceso a una parte considerable de las pruebas aplicadas pese a los múltiples intentos de comunicación realizados con los encargados de las evaluaciones en las comunidades autónomas. La segunda limitación, derivada de la primera, atañe al tamaño de la muestra: desconocemos en qué porcentaje la muestra disponible es representativa de la totalidad de los ítems. Resultó igualmente imposible obtener esta información de las Consejerías de Educación aunque se realizaron repetidos intentos. Ambas limitaciones quedaron fuera del control de la investigadora.

CAPÍTULO 4

ANÁLISIS DE LOS ÍTEMS

Este capítulo describe el análisis realizado. Primero, se presenta la clasificación de los ítems en cada Comunidad Autónoma según las variables de estudio. En segundo lugar, se muestran los resultados del análisis logarítmico-lineal realizado para determinar el grado de dependencia o interacción entre las variables. En tercer lugar, se incluye una descripción del análisis realizado a cada uno de los 173 ítems que componen la muestra no sólo según las variables de estudio sino también atendiendo a las competencias potenciales que el estudiante requiere movilizar para completar la tarea. En cuarto lugar, se presenta una descripción de las características comunes que exhiben los ítems. Finalmente, se describen las fortalezas y deficiencias identificadas en los ítems junto con ejemplos de ítems con fortalezas e ítems con deficiencias.

1. Análisis de los ítems por Comunidad Autónoma

De acuerdo con el modelo matemático PISA y el enfoque funcional del conocimiento matemático se analizaron los 173 ítems disponibles en la muestra de cinco pruebas de diagnóstico. Para cada comunidad se evaluaron los ítems de acuerdo a las variables de tarea consideradas en el estudio: contexto, nivel de complejidad y contenido. Primero se analizaron los ítems de acuerdo al contexto. La Tabla 2 muestra la distribución de las frecuencias de los ítems por comunidad y contexto. La Tabla 3 presenta la distribución porcentual correspondiente.

Tabla 2
Distribución de los ítems por Comunidad Autónoma y contexto

Contexto	Comunidad Autónoma					Total
	A	B	C	D	E	
Personal	5	1	3	13	9	31
Educativo/Laboral	6	16	7	2	0	31
Público	7	58	19	10	5	99
Científico	0	9	3	0	0	12
Total	18	84	32	25	14	173

Tabla 3
Porcentaje de los ítems por Comunidad Autónoma y contexto

Contexto	Comunidad Autónoma					Total
	A	B	C	D	E	
Personal	27,8	1,2	9,4	52,0	64,3	17,9
Educativo/Laboral	33,3	19,0	21,9	8,0	0,0	17,9
Público	38,9	69,0	59,4	40,0	35,7	57,2
Científico	0,0	10,7	9,4	0,0	0,0	6,9
Total	100,0	100,0	100,0	100,0	100,0	100,0

En la distribución del total de ítems de acuerdo al contexto se observa, globalmente, un desequilibrio hacia los ítems de contexto público, cuyos enunciados representan más del 50% del total. Este resultado se mantiene en todas las comunidades excepto la comunidad E, cuyos ítems se inclinan al contexto personal. El 6,9% del total de ítems se clasificó en la categoría de contexto científico destacando que las pruebas de las comunidades A, D y E no incluyeron ítems para este contexto, como se observa en la Tabla 3. La comunidad E tiene la mayor proporción de ítems en el contexto personal (64,3%) sin incluir ítems en los contextos educativo/laboral y científico. La comunidad A incluyó el mayor número de ítems en el

contexto educativo/laboral (33,3%) y no incluyó ítems en el contexto científico. La comunidad con la mayor proporción de ítems en el contexto público es la comunidad B (69,0%) así como en el contexto científico (10,7%). La comunidad C igualmente incluyó una alta proporción de ítems en el contexto público (59,4%). La figura 4 ilustra estos resultados.

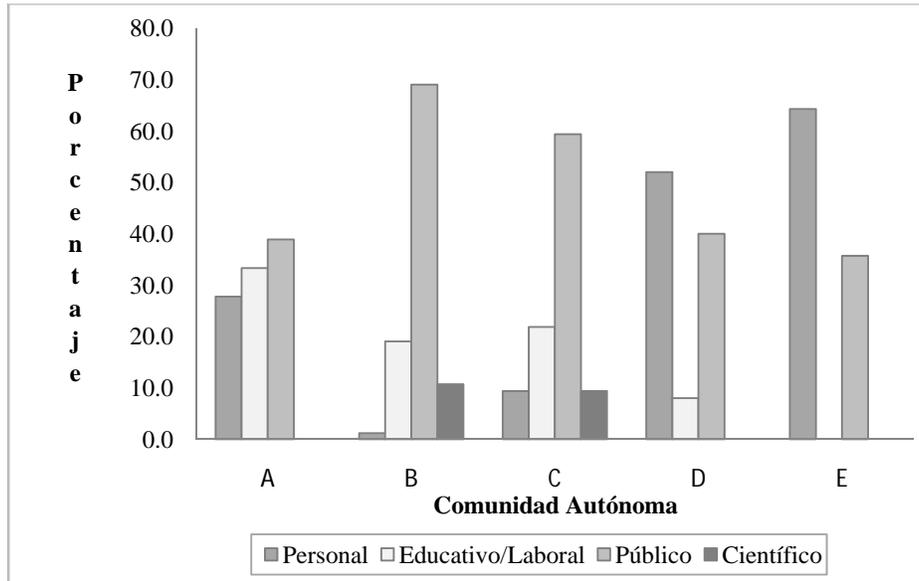


Figura 4. Porcentaje de ítems por contexto y Comunidad Autónoma

La representación gráfica muestra que existe un evidente desequilibrio en la distribución de ítems de acuerdo con esta variable, que difiere entre unas comunidades y otras. El contexto público predomina en las comunidades B y C, mientras que los contextos personal y científico tienen en ambas un porcentaje muy bajo. A su vez, el contexto personal predomina en las comunidades D y E, con porcentajes superiores al 50% de los ítems, mientras que los contextos educativo y científico están prácticamente ausentes en ambas comunidades. De todos los contextos, el científico parece presentar mayor dificultad para elaborar enunciados.

En relación con la variable contexto todas las comunidades muestran limitaciones, que suponen el predominio de un contexto, la ausencia de otro contexto, o bien la conjunción de esas dos limitaciones.

La segunda variable analizada fue el contenido. La distribución de los ítems por comunidad y contenido y la distribución porcentual correspondiente se presentan en las Tablas 4 y 5, respectivamente.

Tabla 4
Distribución de los ítems por Comunidad Autónoma y contenido

Contenido	Comunidad Autónoma					Total
	A	B	C	D	E	
Cantidad	3	18	9	11	3	44
Cambio y relaciones	6	25	2	1	3	37
Espacio y forma	7	21	8	4	4	44
Incertidumbre	2	20	13	9	4	48
Total	18	84	32	25	14	173

Tabla 5
Porcentaje de los ítems por Comunidad Autónoma y contenido

Contenido	Comunidad Autónoma					Total
	A	B	C	D	E	
Cantidad	16,7	21,4	28,1	44,0	21,4	25,4
Cambio y relaciones	33,3	29,8	6,3	4,0	21,4	21,4
Espacio y forma	38,9	25,0	25,0	16,0	28,6	25,4
Incertidumbre	11,1	23,8	40,6	36,0	28,6	27,7
Total	100,0	100,0	100,0	100,0	100,0	100,0

En cuanto al contenido, la distribución del total de ítems en las categorías de contenido es equilibrada, con leves diferencias. La mayor frecuencia se observa en el contenido de incertidumbre (27,7%). Las clasificaciones de cantidad y espacio y forma exhiben igual porcentaje (25,4%). Aunque las diferencias por comunidades son apreciables, globalmente hay equilibrio entre los contenidos utilizados para la totalidad de los ítems. En la distribución de los ítems en cada comunidad de acuerdo al contenido, se aprecia que la comunidad D incluyó la mayor proporción de ítems en el contenido de cantidad (44,0%). En el contenido de cambio y relaciones la comunidad A se adjudicó la delantera con 38,9%. Esta comunidad también incluyó la mayor proporción de ítems en el contenido de espacio y forma (33,3%).

En cuanto al contenido de incertidumbre, fue la comunidad C quien incluyó el mayor número de ítems en este contenido (40,6%). Tanto la comunidad C como la D incluyeron las proporciones más bajas de ítems en el contenido de espacio y forma: 6,3% y 4,0%, respectivamente. En la comunidad B se observa una distribución equilibrada en el contenido de los ítems que, igualmente, se observa en la comunidad E más no así en las otras comunidades. La figura 5 ilustra estos resultados.

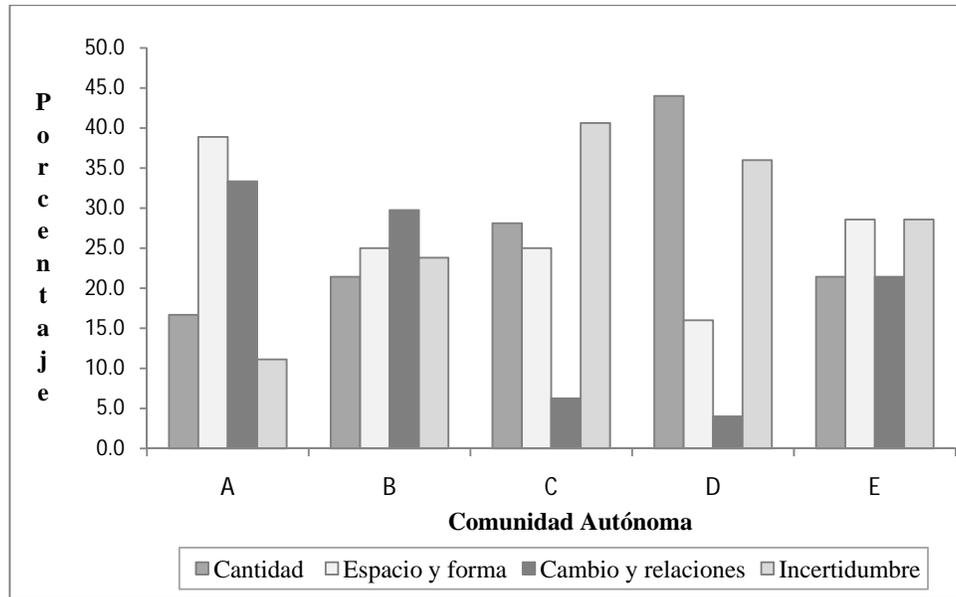


Figura 5. Porcentaje de los ítems por contenido y Comunidad Autónoma

Puede apreciarse en el gráfico anterior que la variable contenido presenta equilibrio global en su distribución de frecuencias, equilibrio que se refleja en las comunidades B y E. También se destaca que ningún contenido está ausente de los enunciados propuestos por las distintas comunidades. No obstante se producen ciertos desequilibrios en algunas comunidades, como es el caso del contenido de cambio y relaciones, ya señalado en las comunidades C y D.

Respecto del nivel de complejidad, la distribución de los ítems en cada Comunidad Autónoma se presenta en la Tabla 6 y en la Tabla 7, la distribución porcentual correspondiente.

Tabla 6
Distribución de los ítems por Comunidad Autónoma y nivel de complejidad

Complejidad	Comunidad Autónoma					Total
	A	B	C	D	E	
Reproducción	5	38	18	15	5	81
Conexión	11	39	11	8	9	78
Reflexión	2	7	3	2	0	14
Total	18	84	32	25	14	173

Tabla 7
Porcentaje de los ítems por Comunidad Autónoma y nivel de complejidad

Complejidad	Comunidad Autónoma					Total
	A	B	C	D	E	
Reproducción	27,8	45,2	56,3	60,0	35,7	46,8
Conexión	61,1	46,4	34,4	32,0	64,3	45,1
Reflexión	11,1	8,3	9,4	8,0	0,0	8,1
Total	100,0	100,0	100,0	100,0	100,0	100,0

Las observaciones respecto del nivel de complejidad arrojan que la mayor frecuencia ocurre en el nivel de reproducción con 46,8% aunque el nivel de conexión (45,1%) es muy cercano. Invariablemente, el nivel de reflexión se aleja considerablemente de los niveles de reproducción y conexión: sólo el 8,1% del total de ítems se clasificó en este nivel. Este resultado se observa de la misma manera en la distribución por comunidades. También en este caso se detecta un sesgo global, debido al bajo porcentaje de ítems de reflexión. En la distribución de los ítems por nivel de complejidad en cada Comunidad Autónoma se observa que las comunidades C y D exhiben la proporción más alta de ítems en el nivel de reproducción: 60,0% y 56,3%, respectivamente. La mayor proporción de ítems en el nivel de conexión se registró en las comunidades E (64,3%) y A (61,1%). La prueba de la comunidad E no incluyó ítems de reflexión. La comunidad B, por su parte, incluyó la mayor proporción de ítems en el nivel de conexión (46,4%) manteniendo una proporción parecida en el nivel de

reproducción (45,2%). La comunidad A incluyó poco más de tres quintas partes (61,1%) de sus ítems en el nivel de conexión. Esta comunidad incluyó mayor proporción de ítems en el nivel de reflexión que las demás comunidades (11,1%). Aunque la proporción más baja de ítems se incluyó en este nivel, sólo la comunidad E no incluyó ítems de reflexión. La Figura 6 representa estos resultados.

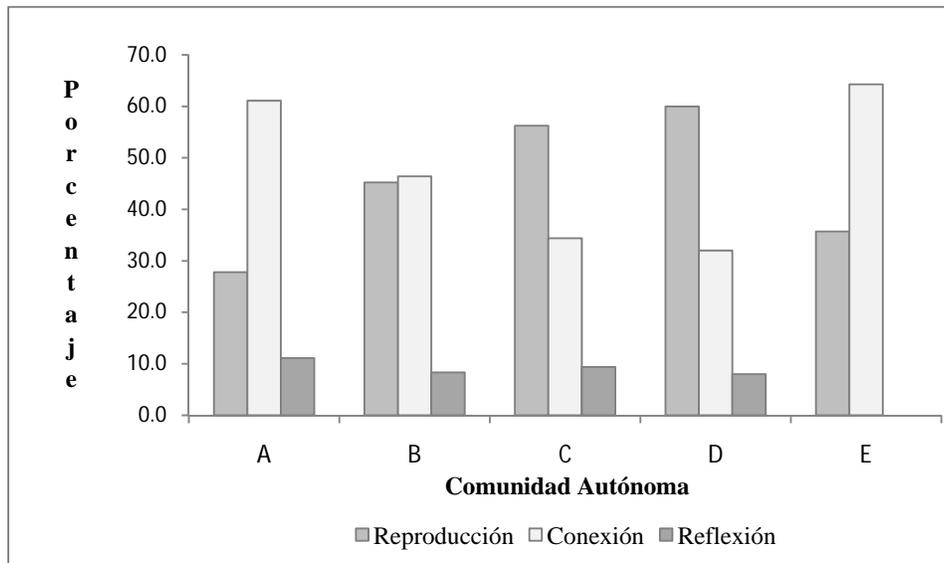


Figura 6. Porcentaje de los ítems por nivel de complejidad y Comunidad Autónoma

Se destaca en la representación gráfica que en las comunidades A, B, C y D se clasificó el menor porcentaje de ítems en el nivel de reflexión mientras que en la comunidad E no se clasificaron ítems en este nivel. Las comunidades A y E incluyeron mayor proporción de ítems en el nivel de conexión mientras que la mayor proporción de ítems en las comunidades C y D se clasificaron en el nivel de reproducción. La comunidad B, por su parte, incluyó proporciones similares de ítems en los niveles de reproducción y conexión.

2. Análisis de los ítems por variables cruzadas

En este apartado se presentan tablas de contingencia para las variables de estudio. Primero se analiza la relación entre las tres variables y luego, entre parejas de ellas. La Tabla 8 presenta el porcentaje correspondiente al cruce de las variables contexto, complejidad y contenido.

Tabla 8
Porcentaje de los ítems por contexto, complejidad y contenido

Contexto	Complejidad	Contenido			Total	
		Cantidad	Cambio y relaciones	Espacio y forma		Incertidumbre
Personal						
	Reproducción	5,2	0,0	0,0	1,2	6,4
	Conexión	0,6	4,0	0,6	5,2	10,4
	Reflexión	0,0	0,0	0,0	1,2	1,2
Educativo/ Laboral						
	Reproducción	1,7	2,0	4,6	1,2	9,5
	Conexión	0,0	5,8	1,7	1,2	8,7
	Reflexión	14,4	0,0	0,0	0,0	14,4
Publico						
	Reproducción	13,9	0,0	11,0	5,2	30,1
	Conexión	0,0	9,2	6,4	7,5	23,1
	Reflexión	1,2	0,6	1,2	1,2	4,0
Científico						
	Reproducción	1,7	0,0	0,0	0,0	1,7
	Conexión	1,2	0,0	0,0	1,7	2,9
	Reflexión	0,0	0,0	0,0	2,3	2,3
TOTAL					100,0	

Se observa en la tabla que la proporción más alta de ítems en el contexto personal ocurre en el contenido de cantidad y el nivel de reproducción (5,2%) con igual porcentaje en el contenido de incertidumbre y nivel de conexión. En el contexto educativo/laboral la clasificación de los ítems se concentra en el contenido de cantidad y el nivel de reflexión (14,4%). Cabe destacar que este porcentaje representa la totalidad de los ítems en el contexto educativo/laboral, contenido de cantidad y el nivel de reflexión.

Por otro lado, los ítems en el contexto público se concentran en el contenido de cantidad y el nivel de reproducción (13,9%) seguido muy de cerca por los ítems en el nivel de reproducción y el contenido de espacio y forma (11,0%). En general, pocos ítems fueron clasificados en el contexto científico. Para este contexto, similar porcentaje de ítems se

clasificaron en el nivel de reproducción y contenido de cantidad que en el nivel de conexión y contenido de incertidumbre.

La Tabla 9 muestra los porcentajes de los ítems por contenido y nivel de complejidad.

Tabla 9
Porcentaje de los ítems por contenido y nivel de complejidad

Contenido	Nivel de complejidad			Total
	Reproducción	Conexión	Reflexión	
Cantidad	22,5	1,7	1,7	25,9
Cambio y relaciones	1,2	19,1	0,6	20,9
Espacio y forma	15,6	8,7	1,2	25,5
Incertidumbre	7,5	15,6	4,6	27,7
Total	46,8	45,1	8,1	100,0

La frecuencia mayor ocurre en los ítems de cantidad en el nivel de reproducción (22,5%) seguido de los ítems de cambio y relaciones en el nivel de conexión (19,1%). Observamos que la frecuencia menor (0,6%) ocurre en los ítems de cambio y relaciones en el nivel de reflexión. Frecuencias bajas también se observan en los ítems de cambio y relaciones en el nivel de reproducción (1,2%) y en los ítems de espacio y forma en el nivel de reflexión (1,2%).

En la Tabla 10 se presenta el porcentaje de ítems al cruzar las variables complejidad y contexto.

Tabla 10
Porcentaje de los ítems por contexto y nivel de complejidad

Contexto	Nivel de complejidad			Total
	Reproducción	Conexión	Reflexión	
Personal	6,4	10,4	1,2	17,9
Educativo/Laboral	8,7	8,7	0,6	17,9
Público	30,1	23,1	4,0	57,2
Científico	1,7	2,9	2,3	6,9
Total	46,8	45,1	8,1	100,0

El porcentaje más alto se observa en los ítems de contexto público en el nivel de reproducción (30,1%) seguido de los ítems en el mismo contexto y nivel de conexión (23,1%). La proporción más baja se observa en los ítems de contenido educativo o laboral en el nivel de reflexión (0,6%). Asimismo, exhiben frecuencias bajas los ítems clasificados en el contexto personal y nivel de reflexión (1,2%) y los ítems de contexto científico y nivel de reproducción (1,7%).

El análisis de los ítems por las variables contexto y contenido (Tabla 11) revela que ningún ítem en el contexto científico se clasificó en los contenidos de cambio y relaciones y de espacio y forma.

Tabla 11
Porcentaje de los ítems por contenido y contexto

Contenido	Contexto				Total
	Personal	Educativo/Laboral	Público	Científico	
Cantidad	5,8	2,3	15,0	2,9	26,0
Cambio y relaciones	4,0	6,9	9,8	0,0	20,8
Espacio y forma	0,6	6,4	18,5	0,0	25,4
Incertidumbre	7,5	2,3	13,9	4,0	27,8
Total	17,9	17,9	57,2	6,9	100,0

Sólo el 0,6% se clasificó en el contenido de espacio y forma y contexto personal representando así la proporción más baja. La proporción más alta se observa en los ítems del contenido espacio y forma y contexto público (18,5%). Los ítems clasificados de contexto público exhiben la mayor frecuencia en todos los contenidos.

3. Análisis logarítmico lineal: Interacción entre las variables

En el capítulo 3 argumentamos que el análisis logarítmico lineal permite determinar si existe relación entre las variables independientes de tarea. El análisis log lineal, una prueba estadística de dependencia, se aplica en casos como el nuestro, donde existen dos o más variables categóricas o nominales y el investigador interesa saber si existen relaciones de dependencia entre ellas. Usando el programado *Statistical Package for the Social Sciences* (SPSS) se realizó un análisis de este tipo para determinar el grado de interacción o dependencia entre las variables contexto, contenido y complejidad. Se consideró la totalidad de los ítems sin clasificarlos de acuerdo a la Comunidad Autónoma a la cual pertenecían. La tabla de contingencia nº 8 en la página 40 se usó como base para este análisis.

Se aprecia en la tabla de referencia la presencia de numerosas casillas con valor de 0. Para compensar esta ausencia de frecuencias, el SPSS asigna a cada casilla una frecuencia de 0.5 lo cual no altera significativamente el resultado final. La prueba estadística de asociaciones parciales que produjo el modelo utilizado se presenta en la Tabla 12.

Tabla 12
Resultado de las pruebas estadísticas de asociaciones parciales

Nombre del efecto	Grados de libertad	Chi-Cuadrado parcial	Probabilidad
contexto*contenido*complejidad	18	8,836	0,9634
contexto*contenido	9	45,336	0,0000
contexto*complejidad	6	13,640	0,0339
contenido *complejidad	6	84,711	0,0000

Se desprende de la estadística generada que la interacción entre las variables contexto, contenido y complejidad no tiene un efecto significativo. Cuando se analiza el efecto de las variables por parejas, se observa que un efecto significativo ocurre en la interacción entre el contenido y la complejidad así como entre el contexto y el contenido. Un menor valor del efecto de la interacción entre las variables por parejas se observa entre el contexto y la complejidad.

4. Evaluación de los ítems según las variables de tarea

Además de clasificar cada ítem de acuerdo a las variables contexto, contenido y nivel de complejidad, se estudió qué competencias debía eventualmente poner de manifiesto el estudiante para resolver el problema y se realizaron algunas observaciones generales del ítem. Por ser una variable de desempeño y no de tarea, la identificación de las competencias se realizó como un recurso de apoyo para caracterizar el ítem. Enmarcada en el modelo matemático de Pisa, esta caracterización tiene en cuenta ocho tipos, ya descritos anteriormente en el apartado 3.1.3 del capítulo 2 (página 21). A continuación se presentan cuatro ejemplos de ítems evaluados, cada uno de ellos clasificado según un contexto diferente. No obstante, puede apreciarse que, igualmente, se ejemplifican los cuatro contenidos y los tres niveles de complejidad.

4.1. Ítem en el contexto personal

El siguiente ítem se clasificó en el contexto personal porque se involucra al estudiante en la práctica diaria del reciclaje en el hogar. El estímulo inicial fue titulado *Desarrollo sostenible* y consta de cinco preguntas.

La modernización de nuestro país exige que asumamos colectivamente el reto y la oportunidad que ofrece un modelo de desarrollo sostenible que haga compatible una dinámica de prosperidad económica, conjuntamente con el aumento del bienestar social y la mejora del medio ambiente. Separar los distintos tipos de vertidos reciclables (papel, vidrio, plásticos y latas) en los hogares para depositarlos en los correspondientes contenedores es una acción que todos deberíamos hacer.

En un instituto de una localidad oscense se ha realizado una encuesta a los alumnos para conocer si tienen el hábito de reciclar. Al elegir al azar un alumno del instituto en el que hemos hecho la encuesta, si nos dice que no recicla nada, ¿es más probable que sea chico o chica?

- A. Chico.
- B. Chica.
- C. Ambos tienen la misma probabilidad.
- D. El planteamiento del problema es absurdo.

El análisis realizado para este ítem se resume en el cuadro siguiente.

Tabla 13
Análisis de ítem en el contexto personal

Contexto	Contenido	Nivel de complejidad	Competencias	Observaciones
Personal. La responsabilidad de cada ciudadano frente a la práctica del reciclaje	Incertidumbre. El estudiante hace inferencias	Reflexión. La tarea es poco estructurada y nada rutinaria; el estudiante analiza la situación para llegar a conclusiones.	Razonar que la probabilidad es la misma para ambos sexos. Expresar que reciclar no es cuestión de género. Reflexionar para ofrecer argumentos razonados. La tarea moviliza las competencias: comunicar, argumentar, y razonar.	La tarea resulta interesante por su planteamiento. Requiere que el estudiante reflexione el argumento que escoge, que bien pudiera conducir a prejuicios sexistas. El contexto pierde relevancia y es de orden cero. Moviliza distintas competencias.

4.2. Ítem en el contexto educativo/laboral

El ítem que se presenta a continuación se clasificó en el contexto educativo/laboral porque al estudiante se le presenta un ambiente laboral, que se aleja de su realidad. El estímulo inicial fue titulado *Panecillos* y consta sólo de este ítem.

La cadena de bollería “La merienda” hace diariamente 360 panecillos empaquetándolos en bolsas de media docena y de una docena de piezas. El distribuidor reparte las bolsas de panecillos entre las cuatro tiendas de la cadena de la siguiente manera:

En “La merienda 1” deja la tercera parte de los panecillos.

En “La merienda 2” deja la cuarta aparte.

En “La merienda 3” deja la quinta parte.

En “La merienda 4” deja las bolsas que quedan.

¿Cuántos panecillos vende cada una de las cuatro tiendas de la cadena? Expresa tus cálculos.

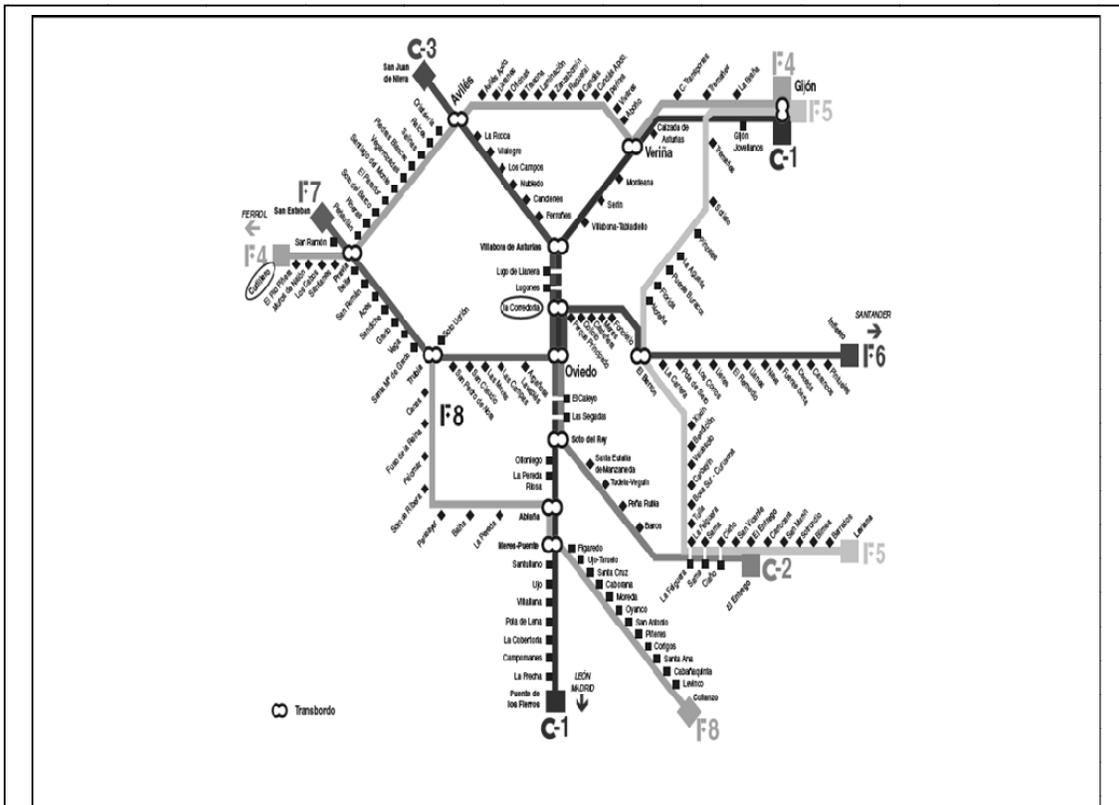
La Tabla 14 resume el análisis realizado para este ítem.

Tabla 14
Análisis de ítem en el contexto educativo/laboral

Contexto	Contenido	Nivel de complejidad	Competencias	Observaciones
Educativo/ Laboral. El problema se refiere a una situación laboral a la cual se enfrenta un comerciante	Cantidad. El estudiante usa la información para calcular la cantidad de panecillos que vende cada tienda.	Reproducción. El estudiante usa una fracción como operador; reproduce un procedimiento conocido que ya ha practicado anteriormente.	Comprender lo que pide el problema. Identificar la relación sobre la que se plantean las preguntas. Comprender la función de la fracción en la resolución del problema. Determinar la operación que debe realizar. La tarea moviliza las competencias: Pensar y razonar, y usar lenguaje simbólico, formal y las operaciones.	La tarea es rutinaria. El contexto no es relevante ni necesario para resolver el problema. Mencionar que se empaquetan los panecillos en bolsas de media y de una docena es información superflua, ya que no es necesaria para resolver el problema y que, incluso, pudiera conducir a confusión. En la tarea se mencionan panecillos y bolsas, pero el problema pide calcular panecillos. Moviliza distintas competencias.

4.3. Ítem en el contexto público

El siguiente ítem se clasificó en el contexto público porque presenta un plano de los trenes de cercanías de cierta ciudad española. Aunque para algunos estudiantes éste pudiera representar un contexto personal, no se clasificó de esta manera por la forma despersonalizada como fue abordado. El estímulo inicial fue titulado *Trenes de cercanías* y consta de siete preguntas.



Indica cuál es el recorrido más corto para trasladarse desde La Corredoria hasta Cudillero realizando el menor número de trasbordos posible.

- A. C1 (desde La Corredoria hasta Oviedo), F1 (desde Oviedo hasta Pravia) y F4 (desde Pravia hasta Cudillero).
- B. C1 (desde La Corredoria hasta Gijón) y F4 (desde Gijón hasta Cudillero)
- C. C3 (desde la Corredoria hasta Avilés) y F4 (desde Avilés hasta Cudillero)
- D. C1 (desde La Corredoria hasta Veriña) y F4 (desde Veriña hasta Cudillero)

La Tabla 15 resume el análisis realizado para este ítem.

Tabla 15
Análisis de ítem en el contexto público

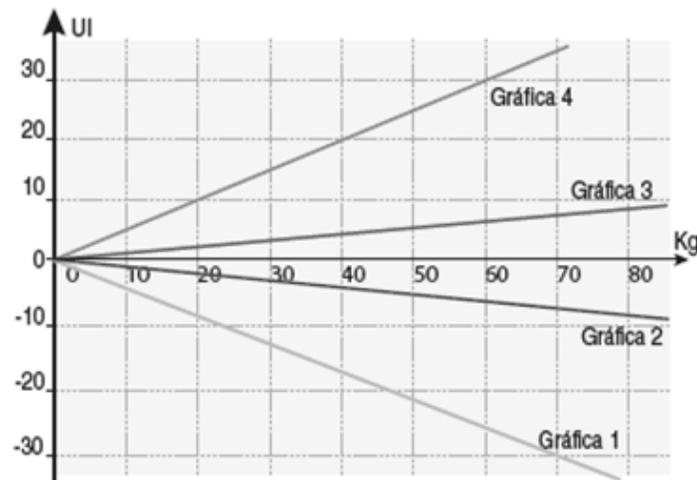
Contexto	Contenido	Nivel de complejidad	Competencias	Observaciones
Público. Se muestra el trayecto de las líneas de trenes de cercanías de una ciudad.	Espacio y forma. La tarea requiere que el estudiante decodifique información visual; el estudiante debe comprender la relación entre el plano de los trenes de cercanías y los trayectos en el entorno real.	Conexión. Aunque la situación le resulte familiar al estudiante, es una tarea no rutinaria. El estudiante debe relacionar la información gráfica con la realidad.	Decodificar información visual. Proveer información específica a partir de un diagrama. Orientarse en el espacio. Relacionar el diagrama con la realidad. Decidir cuál es la mejor respuesta de entre varias opciones. Moviliza las competencias: argumentar, comunicar, modelizar, representar.	El contexto es relevante y necesario para resolver el problema. El estudiante debe moverse del contexto al problema y viceversa por lo cual el contexto se clasifica de segundo orden. La imagen es atractiva, clara y fácil de interpretar. El ítem es similar al problema <i>Sistema de transporte</i> aplicado en PISA 2003. El estudiante requiere desplazarse del contexto al mundo real. Permite la toma de decisiones. Situación relevante. Moviliza distintas competencias.

4.4. Ítem en el contexto científico

El siguiente ítem se clasificó en el contexto científico porque presenta una situación relacionada con la incidencia y prevalencia de la diabetes, un tema con el cual un estudiante de este nivel pudiera no estar familiarizado. El estímulo inicial fue titulado *Dosis fatal* y consta de siete preguntas.

La diabetes es una enfermedad que está aumentando de forma preocupante entre las personas jóvenes debido a los malos hábitos alimenticios y la obesidad. Consiste en que el cuerpo es incapaz de controlar los niveles de glucosa en sangre. El tratamiento más común es inyectar insulina diariamente en dosis adaptadas a cada persona. El cálculo de la dosis debe ser correcto, de lo contrario puede resultar perjudicial.

La dosis diaria de la insulina “Lispro” es de 0,1 Unidades de Insulina (UI) por Kg de peso. ¿Qué gráfica representa la dosis de “Lispro” que hay que administrar por Kg de peso?



- A. Gráfica 1
- B. Gráfica 2
- C. Gráfica 3
- D. Gráfica 4

El siguiente cuadro resume el análisis realizado para este ítem.

Tabla 16
Análisis de ítem en el contexto científico

Contexto	Contenido	Nivel de complejidad	Competencias	Observaciones
Científico. El estudiante está poco relacionado con la incidencia y prevalencia de enfermedades.	Cambio y relaciones. El estudiante interpreta datos presentados en la gráfica para establecer la relación entre las variables.	Conexión. El estudiante reconoce la relación entre el peso y las unidades de insulina.	Comprender la información que se presenta. Interpretar la información para determinar la relación entre las variables. Reconocer que existe una constante de proporcionalidad. Leer la gráfica para determinar la solución. La tarea moviliza las competencias: representar, pensar y razonar, comunicar.	El contexto es relevante. La tarea es no rutinaria. La tarea moviliza distintas competencias. Se permite la toma de decisiones basadas en información relevante.

5. Características comunes de los ítems

Independientemente de la comunidad donde fueron aplicados, los ítems de las cinco pruebas que componen la muestra presentan algunas características comunes. La primera característica común que identificamos es la manera de presentar la situación. La segunda característica se refiere a las posibles limitaciones y fortalezas de los ítems.

5.1. Presentación de la situación

Los ítems incluidos en las pruebas usan diversas maneras para presentar la situación en la cual se enmarca el problema. La distribución de los ítems según la manera de presentar la situación se resume en la Tabla 17.

Tabla 17
Distribución de los ítems por la forma de presentar la situación

Forma de presentar la situación	Frecuencia	Porcentaje
Redacción en párrafo o relato	65	37,5
Gráfico estadístico	43	24,9
Diagrama	32	18,5
Tabla de datos	17	9,8
Plano	14	8,1
Mapa	2	1,2
Total	173	100,0

Es notable la alta proporción de ítems que presentan la situación de la manera más tradicional: redactados en un párrafo o en forma de relato (37,5%). Estos ítems no están acompañados de formas alternas o complementarias de representación que pudieran servir de apoyo a la información provista en el problema. No obstante la forma narrativa en que se presentaron, estos ítems movilizan diversas tareas: calcular cantidades, dibujar una gráfica, derivar una fórmula, hacer inferencias, aplicar una fórmula o completar una tabla.

El segundo recurso más usado es el gráfico. Aproximadamente una cuarta parte de los ítems (24,9%) presenta la situación mediante una gráfica. Este método igualmente se usó para movilizar tareas tales como deducir una ecuación, calcular cantidades, hacer inferencias y explicar el comportamiento de las variables involucradas. Representando cerca de una quinta parte del total de los ítems (18,5%) se observa el recurso de los diagramas usado para movilizar tareas tales como establecer relaciones de proporcionalidad, analizar relaciones geométricas, escribir expresiones algebraicas, calcular cantidades e identificar las características de una figura plana o un cuerpo geométrico.

Los recursos menos usados para presentar las situaciones fueron las tablas de datos (9,8%), los planos (8,1%) y los mapas (1,2%).

5.2. Fortalezas y deficiencias de los ítems

¿Qué hace a un ítem bueno o deficiente? De acuerdo al modelo matemático de PISA y al enfoque funcional del aprendizaje matemático, un problema presenta una situación real que provee un contexto para la tarea. Se impone así una demanda cognitiva con un determinado nivel de complejidad a la cual el estudiante se enfrenta usando como herramientas sus conocimientos matemáticos, que moviliza mediante sus capacidades y que maneja por medio de diversas competencias para completar el proceso de resolución. Dentro de ese marco de referencia calificamos un ítem como *adecuado* si, en términos generales, las tareas que propone cumplen las siguientes condiciones.

- Las situaciones reales que presenta son de cierta relevancia, actual o potencial, para el estudiante.
- La relevancia del contexto se mantiene durante el proceso de resolución.
- La tarea no es rutinaria sino que estimula el razonamiento y la interpretación.
- Promueve que el estudiante se desplace del contexto del problema al mundo matemático necesario para resolverlo. Es decir, faculta al estudiante a integrarse en el proceso de matematización.
- Moviliza potencialmente distintas capacidades.
- Estimula en el estudiante la capacidad de tomar decisiones o seleccionar las estrategias que estime más adecuadas para resolver el problema.
- Posee la extensión necesaria para que el estudiante movilice capacidades de orden superior.

De otro lado y dentro del mismo marco de referencia, un ítem se califica como *inadecuado* si no cumple las condiciones mencionadas y si, además, la tarea que propone atiende estrictamente a los contenidos curriculares y no a las competencias. Puesto que el contexto se destaca como un elemento importante en el modelo matemático de PISA, un ítem se calificará como débil contextualmente si el contexto que presenta es un camuflaje, que sólo sirve para ofrecer la apariencia de estar enmarcado en el mundo real, en cuyo caso el contexto es sólo un pretexto o un accesorio.

Los 173 ítems evaluados en este estudio presentan distintas fortalezas y deficiencias. La tabla 18 resume las fortalezas identificadas en los ítems evaluados, en la tabla 19 se presentan las deficiencias identificadas y en la tabla 20, un resumen de las capacidades potenciales que

moviliza la tarea en cada ítem. Es importante enfatizar que un mismo ítem, por ejemplo, pudiera presentar un contexto relevante pero suprimir la toma de decisiones en cuyo caso el ítem presentaría una fortaleza y una deficiencia simultáneamente. Cabe mencionar que el porcentaje se calculó sobre el total de los 173 ítems.

Tabla 18
Porcentaje de ítems con fortalezas

Fortalezas	Comunidad Autónoma					Total	Porcentaje (%)
	A	B	C	D	E		
Contexto relevante	9	61	13	19	9	111	64,2
Tarea no rutinaria	11	39	14	14	9	87	50,3
Moviliza distintas capacidades	10	47	7	9	7	80	46,2
Tiene relevancia práctica	4	42	4	18	7	75	43,3
Permite la toma de decisiones	10	23	7	5	5	50	28,9
Tarea secuencial	9	19	2	9	8	47	27,2
Hay desplazamiento del contexto al mundo matemático	4	8	0	0	3	15	8,7

Se destaca en la tabla que cerca de dos terceras partes de los ítems evaluados presentan un contexto relevante y necesario para resolver el problema (64,2%) mostrando esta fortaleza con la mayor frecuencia. La mitad de los ítems presentan una tarea que consideramos no rutinaria. Con proporciones cercanas al 50% encontramos tareas enmarcadas en la vida real con relevancia práctica para el estudiante y tareas que movilizan distintas capacidades. Por el contrario, sólo el 8,7% de los ítems permite que el estudiante se desplace del contexto del problema al mundo matemático necesario para resolverlo. Es decir, que apenas una décima parte de los ítems incluidos en las pruebas estimula al estudiante a iniciar el proceso de matematización.

Respecto de las deficiencias identificadas en las pruebas, se observa en la tabla 20 que en la mitad de los ítems la tarea es corta y suprime la toma de decisiones. Muy cerca de este porcentaje encontramos tareas rutinarias, tareas que, aunque enmarcadas en la vida real, carecen de relevancia práctica para el estudiante y tareas que no movilizan capacidades de orden superior.

Tabla 19
Porcentaje de ítems con debilidades

Debilidades	Comunidad Autónoma					Total	Porcentaje (%)
	A	B	C	D	E		
Tarea corta que suprime la toma de decisiones	5	35	19	19	10	88	50,9
Tarea rutinaria	6	30	19	11	11	77	44,5
No moviliza distintas capacidades	5	29	21	14	6	75	43,4
Carece de relevancia práctica	6	27	26	7	7	73	42,2
Atiende contenidos curriculares	8	22	12	14	3	59	34,1
Contexto irrelevante o camuflado	5	15	19	6	2	47	27,2

Se observa además que poco más de una cuarta parte de los ítems presentan tareas cuyo contexto está camuflado o es irrelevante e innecesario para resolver el problema. De otro lado, aproximadamente una tercera parte de los ítems presenta tareas que en lugar de movilizar competencias, atienden contenidos curriculares.

Respecto de las capacidades o competencias que potencialmente movilizan las tareas, se observa en la tabla 21 que representar es la capacidad potencialmente movilizada con mayor frecuencia (41,6%).

Tabla 20
Competencias potenciales que movilizan las tareas de los ítems

Competencias	Comunidad Autónoma					Total	Porcentaje (%)
	A	B	C	D	E		
Representar	7	38	7	13	7	72	41,6
Comunicar	6	35	5	8	7	61	35,3
Plantear y resolver problemas	7	11	3	0	3	24	13,9
Utilizar lenguaje simbólico, formal y técnico y las operaciones	7	20	2	0	2	31	17,9
Pensar y razonar	3	13	4	1	2	23	13,3
Argumentar	2	14	1	3	3	23	13,3
Modelizar	0	7	0	0	0	7	4,0

Se observa que la competencia de comunicar es potencialmente movilizada en el 35,3% de los ítems. La competencia de modelizar pudo haber sido movilizada sólo en el 4,0% de los ítems, representando la competencia con el menor porcentaje. Se incluyen a continuación, ejemplos de ítems con fortalezas e ítems con deficiencias.

5.2.1. Ejemplos de ítems con fortalezas

- *Javier tiene una avería en el baño de su casa y necesita con urgencia un fontanero. Nuria, una compañera de trabajo, le ha dado la referencia de dos empresas de fontanería. Javier ha decidido representar conjuntamente las tarifas de ambas empresas para establecer comparaciones, y esto es lo que ha obtenido:*



Como Javier prevé que la reparación de su baño dure unas 5 horas, se ha decantado por la empresa A. Indica a cuanto asciende su factura y razona si ha hecho la elección más conveniente para su bolsillo.

¿En dónde reside la fortaleza de este ítem? Aunque el ítem se sitúa en un contexto público por la competencia en tarifas de dos empresas de fontanería, la situación guarda cercanía con el estudiante por su cotidianidad lo cual le imprime relevancia práctica. Tanto el contexto como el estímulo inicial mantienen su relevancia para la resolución del problema. El problema moviliza potencialmente las competencias de representar, pensar y razonar, argumentar y comunicar. La tarea no es rutinaria, es secuencial y permite la toma de decisiones.

- *Nos vamos a mudar a un piso más grande. Tenemos que hacer la mudanza desde la Plaza del Centeno hasta la Avenida del Estadio. Para hacer la mudanza podríamos usar varios caminos como ves en la figura.*



El Camino de los Barrios y el Camino del Centro atraviesan la ciudad. En ellos se tarda 1,5 minutos en recorrer cada km. Además, en estos dos caminos hay semáforos como ves en

la figura. Cada semáforo en rojo nos hace estar parados dos minutos. El Camino de Circunvalación rodea la ciudad. No tiene semáforos y se viaja normalmente a una velocidad mayor, tardándose 1 minuto en recorrer cada km.

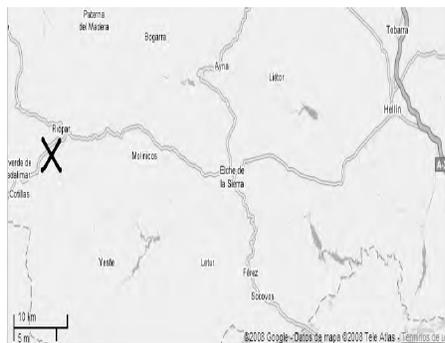
Si queremos elegir el camino que nos suponga menos tiempo, teniendo en cuenta que tendremos que dar muchos viajes, ¿qué camino elegirías? ¿Por qué? Ten en cuenta que no sabemos de antemano si los semáforos estarán en rojo o en verde.

¿Cuáles son, a nuestro juicio, las fortalezas de este ítem? El contexto es relevante y necesario para resolver el problema al mismo tiempo que el estímulo inicial conserva su relevancia durante el proceso de resolución. La tarea demanda que el estudiante se desplace del contexto al conocimiento matemático que pone en juego para resolver el problema. La imagen que se incluye es clara y fácil de interpretar. La tarea es poco estructurada, no es rutinaria y exige razonamiento e interpretación por lo cual moviliza competencias de orden superior entre las que figuran la toma de decisiones y la presentación de argumentos matemáticos que justifiquen su respuesta. Además, la tarea es extensa y exige que el estudiante realice una secuencia de pasos.

- En clase se explica que el agua de la que se abastece la ciudad de Hellín proviene del río que pasa cerca de la localidad. Si la distancia, en línea recta, entre el río y Hellín es, en el mapa, de 12 cm y la escala de 1,9 cm por 10 km.

¿Qué distancia existe entre el nacimiento y la localidad?

- 12 Km.
- 50 Km.
- 63 Km.
- 72 Km.



¿Por qué es adecuado este ítem? Aunque el ítem se desliga del estímulo inicial que consiste en una argumentación respecto de las tendencias en el consumo del agua en los hogares y en la conveniencia de ahorrar agua, el contexto mantiene su relevancia y es necesario para resolver el problema. Además, por su cercanía con la realidad diaria del estudiante, el ítem se

clasifica en el contexto personal. Aunque el estudiante debe recordar y aplicar los conceptos de razón, proporción y factor de escala, no consideramos que la tarea sea rutinaria. La resolución del problema exige realizar una secuencia de tareas para lo cual el estudiante requiere movilizar una serie de competencias como, por ejemplo, pensar y razonar, modelar para moverse entre la realidad y el mundo matemático que presenta el problema, plantear y resolver el problema, representar para usar eficazmente la manera como se presenta la situación y comprender el recurso gráfico que se acompaña. A pesar de que el mapa no es necesario para resolver el problema, sirve de apoyo al ítem.

5.2.2. Ejemplos de ítems con deficiencias

- *El sorteo de lotería más popular en España es el Sorteo Extraordinario de Navidad, que se celebra cada 22 de diciembre. En el sorteo participan 85 000 números. En el año 2008 se emitieron 195 billetes de cada número. Cada billete consta de 10 décimos.*

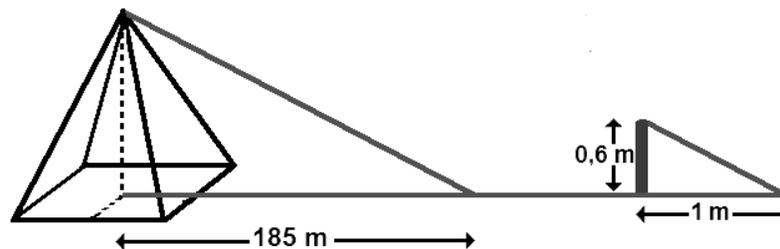


Justifica si la siguiente afirmación relativa al próximo sorteo de lotería de Navidad es correcta: “Según se deduce del gráfico anterior, el cinco es la terminación del gordo que tiene más probabilidad de salir en el próximo sorteo de la lotería de Navidad.”

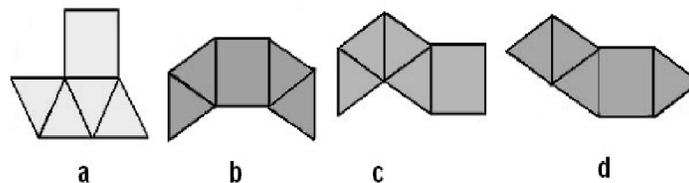
¿Por qué calificamos este ítem como deficiente? Aunque el ítem se clasifica en el nivel de reflexión porque la situación es poco estructurada y exige que el estudiante razone respecto de ella, para hallar la solución sólo debe recordar lo aprendido respecto del concepto de probabilidad por lo que se concluye que la tarea no moviliza competencias ni permite que el estudiante seleccione su propia estrategia de resolución. En este ítem el estímulo inicial es irrelevante, la imagen que se acompaña no cumple ningún objetivo, es accesorio y el contexto está camuflado. La tarea no puede clasificarse como rutinaria pero el estudiante reproduce un procedimiento conocido y previamente practicado. De otro lado, el gráfico al cual se hace referencia se incluye en el ítem anterior lo cual pudiera conducir a confusión además de que pudiera dar la impresión de que el ítem carece de independencia. No obstante, si el estudiante domina el concepto de probabilidad, puede concluir que la afirmación no es correcta, sin necesidad de referirse al gráfico mencionado. Por último, a pesar de que la situación se

enmarca en el mundo real y no le es ajena al estudiante, pudiera carecer de relevancia práctica para él aún cuando los juegos de azar conllevan para cualquier grupo de edad, un interés comúnmente aceptado.

- *En un viaje a Egipto no podemos perdernos la visita a sus pirámides. Observa cómo puede medir la altura de una pirámide. Si colocamos un palo y medimos su sombra, como podemos medir la sombra de la pirámide, basta con relacionar los triángulos rectángulos por semejanza.*



¿Con cuál de los siguientes desarrollos planos se puede construir una pirámide?



¿Cuáles son las deficiencias de este ítem? Si bien es cierto que la tarea exige una alta capacidad de visualización, es igualmente cierto que carece de relevancia práctica para el estudiante, no es un problema de la vida real. El estímulo inicial pierde su relevancia y el contexto es sólo un pretexto. La tarea no es rutinaria pero la situación es ficticia y no tiene conexión con la vida cotidiana. El problema se presenta con fines estrictamente académicos y se refiere a contenidos curriculares específicos. El estudiante decodifica la información del diagrama para determinar el diseño de la pirámide y relaciona un objeto en tres dimensiones con uno en dos dimensiones.

- *El día de la fiesta del Centro, grabamos las actividades y queremos elaborar un DVD con una selección de vídeos y fotos para que los participantes de nuestro curso tengamos un recuerdo.*

Un paquete de 75 DVD nos cuesta 39 euros. Cada DVD nos cuesta:

- A. 0,49 euros*
- B. 0,50 euros*
- C. 0,51 euros*
- D. 0,52 euros*

En este ítem identificamos las siguientes deficiencias: para resolver el problema el estudiante reproduce cálculos algorítmicos por lo que no moviliza competencias; la tarea es rutinaria y suprime la toma de decisiones; aunque la situación le resulte familiar al estudiante, el contexto está camuflado y es innecesario e irrelevante para llegar a la solución; la tarea es corta, trivial y estrictamente matemática; el estímulo inicial no cumple ningún propósito en la resolución del problema.

En el próximo capítulo presentamos un resumen de los principales hallazgos del análisis realizado.

CAPÍTULO 5

EVALUACIÓN DE LOS RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

En esta investigación se aborda la primera aplicación de las evaluaciones autonómicas de diagnóstico en competencia matemática en el segundo curso de la Educación Secundaria Obligatoria realizadas en el término académico 2008-2009. Los resultados de este estudio están basados en el análisis de los ítems incluidos en estas pruebas. Como se mencionara en el capítulo 3, la muestra incluida en esta investigación estuvo constituida por las pruebas aplicadas en cinco Comunidades Autónomas en el año académico bajo estudio. Este capítulo consiste de una evaluación de los resultados presentados en el capítulo 4. En primer lugar, se presenta un resumen de los hallazgos. En segundo lugar, se discuten los resultados del estudio a la luz de los objetivos planteados en el capítulo primero. Finalmente, se presentan las conclusiones generales derivadas de los resultados y se recomiendan algunas posibles investigaciones futuras.

1. Resumen de los hallazgos

Se analizaron los 173 ítems disponibles en la muestra de cinco pruebas de diagnóstico de acuerdo con el modelo matemático de PISA y el enfoque funcional del conocimiento matemático. Primero, fueron clasificados los ítems de cada comunidad para cada una de las variables de estudio: contexto, nivel de complejidad y contenido y estudiadas sus distribuciones de frecuencias. Luego, se realizaron tablas de contingencia para las tres variables y para parejas de variables, acompañadas del análisis logarítmico-lineal correspondiente. Finalmente, se identificaron, de forma general, las fortalezas y debilidades más destacadas de los ítems. A continuación se resumen los principales hallazgos del análisis realizado.

1. Excepto la comunidad A, todas las comunidades muestran desequilibrio en la distribución de los ítems por contexto. Las comunidades D y E no incluyen ítems en el contexto educativo/laboral. Esta última tampoco incluye ítems en el contexto científico.

2. Más de la mitad del total de los ítems (57,2 %) se sitúan en el contexto público mientras que sólo 6,9 % de ellos se clasifican en el contexto científico. Este es un primer sesgo detectado en la redacción de los ítems.
3. La distribución del total de los ítems por contenido se muestra equilibrada con la proporción más alta observada en el contenido de incertidumbre (27.7 %) y la más baja (21.4 %) en el contenido de espacio y forma. No obstante, la comunidad B incluye la proporción más alta de ítems en esta categoría. No se aprecia sesgo en la redacción de los ítems en relación con su contenido.
4. El total de los ítems se distribuyen de manera similar en los niveles de complejidad de reproducción (46,8 %) y conexión (45,1 %). Invariablemente, el nivel de reflexión presenta la menor proporción de ítems tanto en el total como en cada una de las comunidades. La comunidad E no incluye ítems en esta categoría. Este es un segundo sesgo detectado en la redacción de los ítems.
5. Al cruzar las variables se encontró que la proporción más alta de ítems se clasifica en el contexto educativo/laboral, contenido de cantidad y nivel de complejidad de reflexión (14,4 %). Con una proporción de 13,9 % se registran los ítems en el contexto público, contenido de cantidad y nivel de reproducción.
6. Los ítems totales por contenido y nivel de complejidad se distribuyen de manera equilibrada. La proporción más alta se observa en el contenido de espacio y forma y nivel de conexión (19,1 %) mientras que sólo el 0,6 % de los ítems se clasifica en el contenido de cambio y relaciones y nivel de reflexión.
7. Una proporción sustancialmente alta de los ítems se observa en el contexto público y nivel de complejidad de reproducción (30,1 %). Sólo el 0,6 % del total de ítems se clasificó en el contexto educativo/laboral y nivel de reflexión.
8. Igualmente equilibrados se observan los ítems por contexto y contenido. La proporción más alta de ítems se concentra en el contexto público y el contenido de espacio y forma (18,5%). No se clasifican ítems en el contexto científico en los contenidos cambio y relaciones y espacio y forma.
9. El análisis logarítmico-lineal comprueba que no existe interacción significativa entre las tres variables en conjunto. El efecto más significativo ocurre entre el contenido y la complejidad seguido de la interacción entre el contexto y el contenido.

10. La forma predominante de presentar la situación fue mediante la redacción de un párrafo o relato (37,5%). Sólo en el 1,2 % de los ítems se presenta la situación mediante un mapa.
11. Los ítems presentan tanto fortalezas como deficiencias. Las fortalezas identificadas con mayor frecuencia fueron el contexto relevante y necesario para resolver el problema. (64,2%). Con la menor proporción aparece la fortaleza de que el estudiante tuviera que desplazarse del contexto del problema al mundo matemático necesario para resolverlo (8,7%).
12. La deficiencia identificada con mayor frecuencia fue tarea corta que suprime la toma de decisiones (50,9%) mientras que en menor proporción se registran los ítems cuyo contexto es irrelevante o camuflado (27,2%).
13. La capacidad que potencialmente con mayor frecuencia movilizan las tareas fue la representación (41,6%). Por su parte, la tarea potencialmente movilizada con menor frecuencia fue la modelización (4,0%).

2. Evaluación de los resultados

Luego de analizar los resultados, la pregunta esencial que debemos formular en esta etapa de la investigación es qué implicaciones tienen los resultados obtenidos para los objetivos del estudio. En otras palabras, ¿hasta qué punto se cumplieron los objetivos propuestos? Primero, repasamos la conciliación de los objetivos específicos; luego, evaluamos el cumplimiento del objetivo general que orientaba esta investigación.

2.1. Evaluación del objetivo 1

El primer objetivo específico de esta investigación pretendía localizar las pruebas de diagnóstico en competencia matemática aplicadas por las comunidades autónomas a los estudiantes del segundo curso de la Educación secundaria Obligatoria en el año académico 2008-2009. Las pruebas fueron localizadas y, según se resume en el capítulo 3, se logró acceso a ellas según la disposición para ello de los encargados de las evaluaciones de diagnóstico de las Comunidades Autónomas que, en efecto, aplicaron prueba en competencia matemática en el año académico bajo consideración. Debido a que no habían sido liberadas para su divulgación, las pruebas de otras cinco comunidades no pudieron ser accedidas. Hubo dos comunidades que, pese a varios intentos de comunicación, no respondieron a la solicitud de la investigadora y, por lo tanto, se desconoce si aplicaron la prueba correspondiente.

Mediante el acceso a varias fuentes de información determinamos que cinco comunidades o bien no ofrecieron prueba en competencia matemática o no aplicaron prueba en el segundo curso de la ESO. A base de estos resultados, concluimos que el objetivo primero de este estudio fue cumplido.

2.2. Evaluación del objetivo 2

Identificar los valores de las variables de estudio en cada uno de los ítems de acuerdo a la caracterización de las pruebas PISA constituye el segundo objetivo específico de esta investigación. Los 173 ítems se clasificaron para cada comunidad de acuerdo a las variables de estudio: contexto, contenido y nivel de complejidad. Es decir, en cada uno de los ítems se identificaron los valores que exhibían las variables de tarea según definidas en el modelo matemático de PISA. Estos resultados fueron organizados y presentados en tablas de frecuencias y tablas de contingencia. A base de estos resultados concluimos que el segundo objetivo fue alcanzado.

2.3. Evaluación del objetivo 3

El tercer objetivo de este estudio consistía en analizar los ítems de las pruebas de acuerdo a sus características y su ajuste a las evaluaciones PISA. La totalidad de ítems disponibles en la muestra de cinco pruebas de diagnóstico fueron analizados de acuerdo con el modelo matemático PISA y el enfoque funcional del conocimiento matemático. Cada ítem fue analizado no sólo en su ajuste a las variables de tarea del modelo matemático PISA sino también en sus fortalezas, debilidades y las posibles capacidades que requiere movilizar el estudiante para completar cada tarea. La realización de este análisis, satisface el cumplimiento del objetivo tercero.

2.4. Evaluación del objetivo general

El objetivo general que orientaba esta investigación era determinar el grado de ajuste al modelo PISA de los instrumentos elaborados por las comunidades autónomas para atender a la evaluación de diagnóstico en segundo curso de la educación obligatoria en el año académico 2008-2009. En términos generales, se encontró que los ítems diseñados responden a la caracterización del modelo matemático de PISA: se enmarcan en un contexto, atienden un contenido matemático y suponen un nivel de complejidad. No obstante, los ítems evaluados presentan distintas y variadas gradaciones respecto de las categorías de estas tres

variables de tarea. Concluimos así que la evaluación del objetivo general es afirmativa; el objetivo general ha sido cumplido. No obstante, las pruebas presentaban sesgos y deficiencias que llevan a aseverar que no exhibían la capacidad adecuada para satisfacer el propósito para el cual fueron diseñadas.

3. Conclusiones y recomendaciones

Con base en los objetivos que guiaron esta investigación, el marco de referencia que la sostiene y los hallazgos presentados en el capítulo 4, enunciaremos las siguientes conclusiones.

- Existen claras diferencias entre los responsables educativos de las comunidades autónomas respecto al acceso público a las pruebas aplicadas. Hay limitaciones en la transparencia del proceso y una cierta desconfianza hacia los profesionales de la educación, en particular, y los ciudadanos en general.
- Los valores identificados en los ítems para las variables de tarea según definidas en el modelo matemático de las evaluaciones de PISA demuestran que en la redacción de los ítems existe sesgo que favorece el contexto público y el nivel de complejidad de reproducción.
- Los ítems redactados por las Comunidades Autónomas incluidas en la muestra mantienen un equilibrio en lo que a contenido se refiere.
- Puesto que en cada ítem se identificó un contexto, un contenido y un nivel de complejidad, variables de tarea definidas por el modelo matemático de PISA, los ítems aplicados se ajustan al modelo matemático de las evaluaciones de PISA.
- Por los sesgos y las deficiencias identificadas en ellas, las pruebas elaboradas por las Comunidades Autónomas para atender las evaluaciones de diagnóstico presentan limitaciones para satisfacer el propósito para el cual fueron diseñadas.
- Para cumplir con el grado de ajuste adecuado a las evaluaciones PISA, se requiere que las Comunidades Autónomas reevalúen el diseño de las pruebas a la luz de las variables definidas en la caracterización de dichas pruebas.

A la luz de los resultados evaluados, las conclusiones derivadas de estos y la revisión de las áreas que estimamos más apropiadas para la investigación continua sobre las evaluaciones de diagnóstico, sugerimos las siguientes líneas futuras de investigación:

- Un estudio de caso de algunas Comunidades Autónomas con el propósito de determinar su efectividad en el proceso de establecer las metas, la relevancia, la interpretación y el diseño de las pruebas de diagnóstico.
- Un estudio descriptivo de los ítems en las evaluaciones generales de diagnóstico que incluya un contraste con los ítems de las evaluaciones autonómicas de diagnóstico que analizamos en esta investigación.
- Un estudio empírico que incluya el diseño de una prueba que se ajuste al modelo matemático de las evaluaciones de PISA y su aplicación a una población estudiantil con características similares.

REFERENCIAS

- Baker, F. B. (2001). *The basics of Item Response Theory*. College Park: ERIC Clearinghouse on Assessment and Evaluation.
- Boulmetis, J. & Dutwin, P. (2000). *The ABCs of Evaluation*. San Francisco: Jossey-Bass Publishers.
- Cortada de Kohan, N. Teoría de Respuesta al Ítem. *Evaluar*, 2004, (4), 95-110. Consultado en <http://www.revistaevaluar.com.ar>
- Curtis, D. & Boman, P. X-ray your data with Rasch. *International Education Journal*, 2007, 8(2), 249-259. Consultado en <http://iej.com.au>
- De Lange, J. (2006). Mathematical Literacy for Living from OECD-PISA Perspective. *Tsukuba Journal of Educational Study in Mathematics*. Vol. 25, 2006, 13-35. Consultado en www.oecd.org
- Hambleton, R.K. & Hariharan S. (1985). *Item Response Theory: Principles and Applications*; Boston: Kluwer-Nijhoff Publishing.
- Hammill, D. y Usala, P. (2002). *Why Analyze Items. U.S. Immigration and Naturalization Service. IPMAAC Panel Discussion: Developing Defensible Written Test Questions: Art, Science, and Some Guidelines*. Consultado en <http://www.ipacweb.org/conf/02/hamill.pdf>
- INECSE (2005). *PISA 2003. Pruebas de Matemáticas y de Solución de Problemas*. Madrid: Ministerio de Educación y Ciencia.
- Instituto de Evaluación (2006). PISA en detalle: Características del estudio PISA 2003. *Apuntes del Instituto de Evaluación*, (4), Madrid: Ministerio de Educación.
- Instituto de Evaluación (2009). *Evaluación General de Diagnostico 2009: Marco de la Evaluación*. Madrid: Ministerio de Educación.
- Lord, F.M. (1990). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates.
- Lupiáñez, J. L. (2009). *Expectativas de aprendizaje y planificación curricular en un programa de formación inicial de profesores de Matemáticas de secundaria*. Tesis doctoral. Universidad de Granada.
- Lupiáñez, J. L. (2010). *Diseño y selección de tareas para el desarrollo de la competencia matemática*. Trabajo presentado en Matemáticas y Competencias Básicas, Oviedo.

- Maestro, C. (2005). *Las evaluaciones nacionales e internacionales. La mejora de la calidad del sistema educativo: el éxito de todos los alumnos como objetivo*. Trabajo presentado en Seminario de Primavera 2005: La enseñanza de las Matemáticas y el Informe PISA. Madrid: Fundación Santillana. Consultado en www.fundacionsantillana.com
- Ministerio de Educación y Ciencia (2006). Ley Orgánica de Educación. Boletín Oficial del Estado, BOE núm. 106. Madrid: Autores. Disponible en www.educacion.es
- Muñiz, J. & Fidalgo, A. M. (2005). *Análisis de los ítems*. Madrid: La Muralla, S.A.
- OCDE (2005). Manual de análisis de datos de PISA 2003: Usuarios de SPSS® Madrid: Instituto Nacional de Evaluación y Calidad del Sistema Educativo (INECSE).
- OECD (2004). Learning for Tomorrow's World: First Results from PISA 2003. Disponible en www.oecd.org
- OECD (2009). Learning Mathematics for Life: A Perspective from PISA. Disponible en www.oecd.org
- Polya, G. (1962), *Mathematical Discovery: On Understanding, Learning and Teaching Problem Solving*. New York: Wiley.
- Rico, L. (1997). *Consideraciones sobre el currículo de matemáticas para educación secundaria*. En L. Rico (Coord.), E. Castro, M. Coriat, A. Marín, L. Puig, et. Al., La educación matemática en la enseñanza secundaria (pp. 15-38). *Barcelona: ice-Horsori*.
- Rico, L. (2007). La competencia matemática en PISA. *PNA*, 1(2), 47-66.
- Rico, L. (2006). Marco teórico de evaluación en PISA sobre matemáticas y resolución de problemas. *Revista de Educación*, (Extraordinario 2006), 275-294
- Rico, L. & Lupiáñez, J. L. (2008). *Competencias matemáticas desde una perspectiva curricular*. Madrid: Alianza Editor.

ENLACES CONSULTADOS PARA ACCEDER A LAS PRUEBAS DE EVALUACIÓN DE DIAGNÓSTICO

- Consejería de Educación de Andalucía. www.juntadeandalucia.es/educacion/
- Consejería de Educación de Asturias. www.educastur.es/
- Consejería de Educación de Aragón. www.educaragon.org/
- Consejería de Educación de Castilla-La Mancha. www.educa.jccm.es/educa-jccm/cm
- Consejería de Educación de Navarra. www.educacion.navarra.es/portal/

ÍNDICE DE ANEXOS

En el CD que acompaña este documento, se incluyen los siguientes anexos:

Anexo A. Análisis de ítems de Andalucía

Anexo B. Análisis de ítems de Aragón

Anexo C. Análisis de ítems de Asturias

Anexo D. Análisis de ítems de Castilla-La Mancha

Anexo E. Análisis de ítems de Navarra

Anexo F. Análisis logarítmico-lineal