Documentos técnicos

Manual técnico del Plan Nacional para la Evaluación de los Aprendizajes PLANEA 2015.

Educación media superior

Documentos rectores

Instituto Nacional para la Evaluación de la Educación
México

Manual técnico del Plan Nacional para la Evaluación de los Aprendizajes
PLANEA 2015. Educación media superior
Primera edición, 2019

D. R. © Instituto Nacional para la Evaluación de la Educación
Barranca del Muerto 341, Col. San José Insurgentes,
Del. Benito Juárez, C. P. 03900, Ciudad de México

La elaboración de este documento estuvo a cargo de la Unidad de Evaluación del Sistema Educativo Nacional.
El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del INEE. Se autoriza su reproducción por cualquier sistema mecánico o electrónico, para fines no comerciales. Cítese de la siguiente manera:

INEE (2019). *Manual técnico del Plan Nacional para la Evaluación de los Aprendizajes PLANEA 2015. Educación media superior.* México: autor.

PRESENTACIÓN **5**

INTRODUCCIÓN **7**

1. ANTECEDENTES DEL INSTRUMENTO DE EVALUACIÓN **10**

1.1 Marco legal o normativo 10
1.2 La educación media superior en México 11
1.2.1 Principales organismos *14*
1.3 Enlace Media Superior 15
1.4 Características y modalidades de Planea 16
1.5 Planea Media Superior (2015-2016) 17

2. CONCEPTUALIZACIÓN DEL INSTRUMENTO DE EVALUACIÓN **18**

2.1. Planeación general del instrumento de evaluación 18
2.1.1. Propósito de Planea ems *19*
2.1.2. Población objetivo *19*
2.1.3. Condiciones de factibilidad de la prueba *20*
2.1.4. Perfiles de los integrantes de los consejos y comités *21*
2.1.5. Usos previstos *22*
2.2. Diseño del instrumento de evaluación 23
2.2.1. Marco de referencia para el desarrollo del instrumento *23*
2.2.2. Delimitación conceptual del objeto de medida y contenidos a evaluar *25*
2.2.3. Ficha técnica de la prueba *26*
2.3. Comité Académico de Determinación de Contenidos e indicadores de aprendizaje de Planea ems 2017 27
2.3.1. Procedimiento metodológico para el desarrollo del Comité Académico *28*
2.4. Tabla de contenidos e indicadores de aprendizaje 31
2.5. Comité de Especificaciones 33
2.5.1. Procedimiento metodológico para el desarrollo de las especificaciones *33*
2.6. Ejemplos de especificaciones 37

3. DESARROLLO DEL INSTRUMENTO DE EVALUACIÓN **42**

3.1. Comité de Elaboración de Reactivos 42
3.1.1. Procedimiento metodológico para la elaboración de reactivos *43*
3.2. Ejemplos de reactivo 49
3.3. Edición del reactivo 58
3.3.1. Revisión externa del reactivo *58*
3.4. Comité de Validez y Sesgo 59
3.4.1. Procedimiento metodológico para el Comité de Validez y Sesgo *60*
3.5. Prueba piloto 66
3.5.1. Preparativos e integración de la prueba piloto *67*
3.5.2. Revisión de la versión piloto de la prueba *67*
3.5.3. Envío a imprenta para generar cuadernillos *68*
3.5.4. Revisión de los cuadernillos y autorización para su reproducción *68*
3.6. Aplicación de la prueba piloto 68
3.6.1. Determinación de la muestra para el piloteo *69*
3.6.2. Diseño de la aplicación piloto *69*
3.6.3. Protocolo de seguridad para la administración de la prueba piloto *71*
3.6.4. Protocolo para el resguardo de la información *71*
3.7. Estadísticos de los reactivos piloteados 72
3.7.1. Análisis de resultados de los reactivos de la prueba piloto *73*
3.7.2. Análisis estadístico de los reactivos *74*
3.8. Integración de la prueba definitiva 80
3.8.1. Selección de los reactivos de la prueba a partir de los criterios estadísticos *80*

3.8.2. Criterios de equivalencia y distribución de los bloques ... *81*
3.8.3. Revisión de patrones de respuesta correcta ... *82*
3.9. Edición, armado e impresión de las pruebas ... 82
3.9.1. Revisión de las impresiones de la prueba ... *83*
3.9.2. Revisión de la versión definitiva de la prueba ... *83*
3.9.3. Envío a prensa para generar cuadernillos ... *83*
3.9.4. Revisión de los cuadernillos y autorización para su reproducción ... *83*

4. ADMINISTRACIÓN Y RESGUARDO DEL INSTRUMENTO DE EVALUACIÓN ... 84

4.1. Aplicación de la prueba nacional ... 84
4.1.1. Determinación de la muestra matricial ... *85*
4.1.2. Diseño de la aplicación nacional ... *87*
4.1.4. Protocolo para el resguardo de la información ... *93*

5. ANÁLISIS DE RESULTADOS DE INSTRUMENTOS DE EVALUACIÓN ... 95

5.1. Calibración ... 95
5.2. Limpieza y validación de datos ... 96
5.3. Análisis estadístico ... 97
5.3.1. Modelo de puntuación de las respuestas ... *97*
5.3.2. Reporte preliminar del análisis ... *98*
5.4. Reporte final del análisis ... 103
5.5. Modelo de equiparación de resultados en el tiempo ... 107

6. DIFUSIÓN Y USO DE LOS RESULTADOS DEL INSTRUMENTO DE EVALUACIÓN ... 108

6.1. Comité de Niveles de Logro y Puntos de Corte ... 108
6.1.1. Procedimiento metodológico ... *109*
6.1.2. Capacitación de los integrantes en el Método Bookmark ... *112*
6.2. Resultados ... 114
6.3. Análisis estadístico de los resultados de la prueba nacional (considerando los puntos de corte de los niveles de logro determinados en el Comité) ... 117
6.4. Interpretación del análisis estadístico ... 118
de los resultados de la prueba nacional ... 118
6.4.1. Reporte de resultados ... *119*
6.4.2. Informe institucional ... *120*
6.4.3. Cuestionarios de contexto ... *121*
6.4.4. Usos de los resultados ... *122*

7. MANTENIMIENTO DEL INSTRUMENTO DE EVALUACIÓN ... 123

7.1. Indicadores de validez ... 123
7.1.1. Análisis de dimensionalidad, confiabilidad y funcionamiento diferencial del instrumento ... *123*
7.2. Construcción del banco de reactivos ... 124
7.3. Manual técnico del instrumento de evaluación ... 125
7.4. Revisiones posteriores a la aplicación definitiva de Planea ems 2017 ... 126

CONCLUSIONES ... 128

BIBLIOGRAFÍA ... 130

GLOSARIO ... 134

ANEXOS ... 143

Anexo I. Perfiles ... 143
Anexo II. Materiales y formatos de trabajo del Comité Académico ... 145
Anexo III. Materiales y formatos de trabajo del Comité de Especificaciones ... 148

Anexo IV. Materiales y formatos de trabajo del Comité de Elaboración de Reactivos 152
Anexo V. Materiales y formatos de trabajo del Comité de Validez y Sesgo 157
Anexo VI. Precisión del diseño muestral 159
Anexo VII. Promedio de las actividades de la aplicación nacional 166
Anexo VIII. Porcentaje de aplicación por planteles-turno Planea ems 2017 167
Anexo IX. Porcentaje de respuesta en Lenguaje y Comunicación 168
Anexo X. Porcentaje de respuesta en Matemáticas 169
Anexo XI. Porcentaje de respuesta del Cuestionario para el alumno 170
Anexo XII. Escalamiento de logro Planea ems 2017 171
(tratamiento de los datos faltantes en los reactivos de logro) 171
Anexo XIII. Materiales y formatos de trabajo 173
del Comité de Niveles de Logro y Puntos de Corte 173
Anexo XIV. Variables que componen los índices simples 177
y escalas de contexto PLANEA EMS 2017 177

ANEXOS TÉCNICOS 181

Anexo 1. Diseño muestral 181

Anexo 2. Metodología de escalamiento para la Evaluación de Logro referida al Sistema Educativo Nacional (ELSEN) 218

Anexo 3. Reporte técnico de la construcción de índices y escalas de los datos de los Cuestionarios de Contexto 233

Bibliografía 273

Presentación

La educación es un derecho constitucional que está fundamentado en el artículo 3º de la Constitución de los Estados Unidos Mexicanos, así como en diversos documentos internacionales suscritos por México, tales como convenciones, pactos, acuerdos y declaraciones internacionales sobre derechos humanos. La importancia de la educación radica en que repercute directamente en el desarrollo de la equidad social, cultural y económica de la población que se encuentra en edad escolar, además de que constituye un medio integrador que impacta en el mejoramiento de la calidad de vida de los ciudadanos en general. Por ello, la evaluación de la educación es un instrumento necesario para apoyar el derecho de los mexicanos a gozar de una educación de calidad, pertinente y equitativa, así como luchar contra el rezago educativo en las zonas vulnerables del país, ya que a partir de los resultados de las evaluaciones es posible establecer acciones y estrategias para fortalecer el sistema educativo mexicano.

Desde los años noventa, México ha transitado por una ruta de fortalecimiento en el desarrollo de evaluaciones de logro con alcance nacional, principalmente para conocer el estado de la educación en México. Ejemplo de lo anterior son las pruebas de Estándares Nacionales, los Exámenes de la Calidad y el Logro Educativos (EXCALE) y la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE). Así también, ha participado en evaluaciones de índole internacional como el Programa para la Evaluación Internacional de Alumnos (PISA), Estudio Internacional de Educación Cívica y Ciudadana (ICCS) y el Estudio Regional Comparativo y Explicativo del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).

Por mandato constitucional, y en su calidad de organismo público autónomo, el Instituto Nacional para la Evaluación de la Educación (INEE) se ocupa de realizar diversas acciones para evaluar la calidad de la educación obligatoria en México y difundir sus resultados. Con este propósito se diseñan y aplican evaluaciones de los componentes, de los procesos y de los resultados en la educación básica y media superior, tanto pública como privada, con el objetivo de ofrecer información que permita tomar decisiones dirigidas a garantizar el derecho a una educación de calidad.

El artículo 25 de la Ley del Instituto Nacional para la Evaluación de la Educación establece que el INEE tiene como atribuciones primordiales "evaluar la calidad, el desempeño y los resultados del Sistema Educativo Nacional en lo que se refiere a la educación básica y a la educación media superior, tanto pública como privada, en todas sus modalidades y servicios" (INEE, 2013: 18). Asimismo, "el Instituto diseñará y realizará mediciones y evaluaciones que correspondan a componentes, proceso o resultados del Sistema Educativo Nacional respecto a los atributos de educandos, docentes y Autoridades Escolares, así como de las características de instituciones, políticas y programas educativos" (INEE, 2013: 18), encaminados a mejorar la calidad y equidad educativas.

Una de las fuentes principales de información son las evaluaciones de los aprendizajes clave de los estudiantes que, a partir del ciclo escolar 2014-2015 y en coordinación con la Secretaría de Educación Pública (SEP), se llevan a cabo mediante el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA).

PLANEA está integrado por un conjunto de pruebas diseñadas para conocer el dominio que los estudiantes tienen de una serie de aprendizajes clave, con base en los referentes curriculares, al término de la educación preescolar, primaria, secundaria y media superior.

En 2017, por primera vez, el INEE tuvo bajo su responsabilidad el diseño, la aplicación y la difusión de los resultados de la prueba de educación media superior de PLANEA. El presente Manual técnico tiene como propósito describir detalladamente la metodología, los criterios, el diseño matricial y el proceso de construcción de dicha prueba, así como el desarrollo y aplicación de la misma, el análisis, la emisión de resultados y las recomendaciones para su uso, en congruencia con los Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación (2017), determinados por el INEE, sin dejar de observar una metodología acorde con las exigencias y los objetivos particulares de la prueba.

Con esta publicación, el INEE cumple con el estatuto de transparencia y rendición de cuentas que señala su Código de Ética y Conducta, y responde al compromiso de evidenciar que la información arrojada por la prueba es tangible y verificable, ya que se obtuvo a partir del seguimiento de pasos rigurosos y metodológicamente acordes con los lineamientos institucionales y con la investigación educativa.

Introducción

En México, la evaluación del aprendizaje de los estudiantes se lleva a cabo, por un lado, en el aula por parte de los docentes, y por otro lado, a gran escala por instancias externas. Este último enfoque se centra en la aplicación de instrumentos de evaluación a una población representativa de estudiantes, con lo cual se busca determinar su nivel de logro educativo respecto a los aprendizajes alcanzados, al mismo tiempo que apoyar el trabajo que realizan los docentes dentro de los centros escolares. Por ello, la evaluación a gran escala es un instrumento que sirve para valorar en qué grado se cumple el derecho de los mexicanos a recibir una educación de calidad, pertinente y equitativa; además, tiene el propósito de coadyuvar a la mejora de los procesos y resultados educativos, considerando la diversidad del país y construyendo sus referentes con la colaboración de los distintos actores del sistema educativo.

La evaluación a gran escala de los aprendizajes de los alumnos en diferentes momentos de la educación obligatoria se lleva a cabo mediante PLANEA. Este plan se implementó a partir del ciclo escolar 2014-2015 y tiene dos objetivos generales:

> [...] evaluar el desempeño del SEN respecto a los aprendizajes en la educación obligatoria, con distintos grados de desagregación (propósito que corresponde al INEE) y retroalimentar a todas las comunidades escolares del país de [preescolar], primaria, secundaria y educación media superior respecto a los aprendizajes que logran sus estudiantes (propósito que corresponde a la SEP) (INEE, 2018: 10).

Debe considerarse que "PLANEA se organiza a través de dos arreglos metodológicos" (INEE, 2018: 10): el referido al sistema de educación obligatoria (PLANEA SEN) y el diseñado para las comunidades escolares (PLANEA Escuelas), cada uno con finalidades distintas. En PLANEA SEN:

> Los resultados se reportan en el ámbito nacional, por entidad federativa y por estratos escolares, según los distintos tipos de servicio. Esta desagregación permite identificar, entre otras cosas, en qué medida los estudiantes en México adquieren los aprendizajes esperados y el tamaño de las brechas entre subpoblaciones. La información obtenida es un insumo para el planteamiento de políticas educativas nacionales y estatales, así como para la planeación, la programación y la operación del sistema educativo. De igual modo, se busca que esta modalidad informe a la sociedad sobre el estado que guarda la educación obligatoria en términos del logro de aprendizaje de los estudiantes del SEN (INEE, 2018: 12).

Dentro de PLANEA SEN se desprende la evaluación de educación básica (EB) y la de educación media superior (EMS). Ésta última es el referente del presente Manual técnico y tiene por objetivo evaluar el dominio de aprendizajes esenciales de los alumnos del último grado de la EMS en sus diferentes subsistemas, a partir de dos campos disciplinares: Lenguaje y Comunicación, y Matemáticas. Estas asignaturas constituyen áreas instrumentales para el aprendizaje de otras y son esenciales para la continua adquisición de nuevos aprendizajes, son relevantes para el dominio del campo disciplinar y prevalecen en el tiempo a pesar de los cambios curriculares.

Fue el 2017, el primer año que el INEE diseñó y aplicó la prueba, e informó sobre los resultados de PLANEA EMS. Las aplicaciones del 2014, 2015 y 2016 fueron coordinadas por la SEP. La prueba de 2017 evaluó aprendizajes clave que ayudan a consolidar las competencias genéricas y las competencias

disciplinares básicas, señaladas en el Marco Curricular Común (MCC), documento rector del Sistema Nacional de Bachillerato (SNB). Tales aprendizajes integran las habilidades que el estudiante requerirá para desarrollarse de manera integral en su entorno sociocultural y afectivo; esto como disposición de la Reforma Integral de la Educación Media Superior (RIEMS), la cual señala que entre los obstáculos en la EMS está la falta de interés de la población escolar en los contenidos, debido a su poca pertinencia tanto para resolver las exigencias del mundo actual como para atender las características de la población adolescente.

El presente documento está organizado en siete grandes apartados. Salvo el primero, cada uno de ellos corresponde a una de las etapas establecidas por el INEE para el desarrollo de los instrumentos de evaluación, las cuales se encuentran sistematizadas en el documento *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* (2017).

En el apartado "Antecedentes del instrumento de evaluación", se hace una revisión de la legislación en materia de evaluación educativa en México para observar las normativas, los códigos y los principios reguladores del SEN y de las evaluaciones de aprendizajes ligadas a él, y comprender el marco reglamentario y las responsabilidades legales que guiaron la prueba PLANEA EMS 2017. Se explican los conceptos más significativos del SEN en el nivel medio superior: la SEMS, la RIEMS y el SNB, para observar la evolución que ha tenido. Finalmente se realiza una descripción de los instrumentos utilizados previamente para evaluar el logro académico en la EMS (ENLACE MS y PLANEA EMS 2015-2016), específicamente sus fortalezas y áreas de oportunidad. También se describen las dos modalidades de PLANEA.

En "Conceptualización del instrumento de evaluación", se presentan los propósitos de la prueba PLANEA EMS 2017, su población objetivo, sus condiciones de factibilidad, así como los perfiles de los integrantes de los cuerpos colegiados, el marco de referencia desde el que se desarrolló la prueba y la delimitación conceptual del objeto de medida. En este apartado se profundiza en el trabajo de los cuerpos colegiados: en primer lugar, del Comité Académico de Determinación de Contenidos, ya que en éste se definen los contenidos que serán evaluados; en segundo lugar, del Comité de Especificaciones, para detallar el procedimiento metodológico y los criterios que guiaron la realización técnica de las especificaciones de reactivos.

En el apartado "Desarrollo del instrumento de evaluación", se describe la concreción de los referentes teóricos y conceptuales referidos en la etapa del diseño. Se abordan los procedimientos metodológicos, así como los productos obtenidos por cada uno de los cuerpos colegiados responsables de la construcción de los reactivos, de la revisión técnica, de sesgo cultural y lingüístico de los mismos, para mostrar la calidad técnica, de contenido y las medidas para evitar el sesgo cultural de la prueba. Aquí también se describen los protocolos para el piloteo de la prueba en una muestra representativa de la población objetivo, los análisis estadísticos y, finalmente, se detalla el ensamble de la prueba para definir las versiones finales.

En "Administración y resguardo del instrumento de evaluación", se detallan los protocolos de administración de la aplicación definitiva de PLANEA EMS 2017: las medidas y condiciones de seguridad y confidencialidad que se implementaron para dotar de validez al proceso, a los resultados y a los análisis que derivaron de ellos. De igual manera, se detallan los protocolos e inventarios para el resguardo de la información, garantizando una vez más su confidencialidad y rigurosidad

En el apartado “Análisis de resultados del instrumento de evaluación”, se expone el procedimiento para el análisis e interpretación de los resultados de PLANEA EMS 2017 y se presentan los análisis estadísticos realizados.

En tanto que el objetivo de la evaluación sólo se completa cuando la información arrojada por la prueba es dada a conocer a los actores interesados en el apartado “Difusión, uso y resguardo de los resultados del instrumento de evaluación”, se detallan las acciones tomadas para difundir amplia y oportunamente los resultados de PLANEA EMS 2017 y se caracterizan las funciones del Comité de Niveles de Logro y Puntos de Corte, su función, los métodos utilizados y los resultados a los que llegó.

Finalmente, en el último apartado, “Mantenimiento del instrumento de evaluación”, se propone un plan de trabajo que permita, en subsecuentes ocasiones, enriquecer la calidad técnica de este instrumento de evaluación e identificar sus fortalezas y áreas de oportunidad. Lo anterior apoya a que el mantenimiento del instrumento de evaluación guarde una estrecha relación entre el alto nivel académico y la cuidadosa revisión metodológica. Disponer de esta información actualizada, confiable y de calidad, es relevante porque potencializa la toma de decisiones en favor de la mejora educativa.

1. Antecedentes del instrumento de evaluación

El Estado mexicano, en su condición de garante de los derechos constitucionalmente establecidos, tiene la obligación de proporcionar la infraestructura y los mecanismos necesarios para asegurar una educación de calidad, que se traduzca en la integración óptima de los aprendizajes relevantes en la vida de los estudiantes. Este compromiso se refleja en la articulación de normas y estructuras que aseguran la eficiencia del sistema educativo, por lo que el Sistema Nacional de Evaluación Educativa (SNEE) actualmente se rige por estrategias participativas, en las cuales agrupa diferentes sistemas, independientes entre sí, pero tendientes a un mismo fin: la calidad y equidad educativas.

Para comprender cómo se inserta PLANEA EMS dentro de esta estructura, cómo responde a la legislación educativa vigente y a partir de qué marcos se seleccionaron los criterios para su construcción, a continuación se presentan los antecedentes que incidieron en el desarrollo de la prueba.

1.1 Marco legal o normativo

El Título Primero de la Constitución Política de los Estados Unidos Mexicanos contiene, dentro de su Capítulo I, el artículo 3º, que garantiza el derecho a la educación bajo tres lineamientos: que sea obligatoria, gratuita y laica. Además, en su fracción II, inciso d), establece que la educación: "Será de calidad, con base en el mejoramiento constante y el máximo logro académico de los educandos" (DOF, 15 septiembre 2017: 5). Así, para el cumplimiento de este mandato, la fracción IX faculta al Instituto Nacional para la Evaluación de la Educación (INEE) para: "evaluar la calidad, el desempeño y resultados del sistema educativo nacional en la educación preescolar, primaria, secundaria y media superior" (DOF, 15 septiembre 2017: 7) y lo autoriza para diseñar y realizar las evaluaciones de componentes, procesos o resultados del sistema.

Para asegurar el cumplimiento constitucional de proporcionar a la población una educación de calidad, el Estado mexicano ha emitido legislaciones educativas en las que está detallada la forma en que se debe aplicar dichos mandatos. Son cuatro las leyes reglamentarias sobre educación que han existido: la Ley Orgánica de Educación, publicada el 30 de diciembre de 1939; la Nueva Ley Orgánica de la Educación Pública, promulgada el 23 de enero de 1942; la Ley Federal de Educación, decretada el 29 de noviembre de 1973, y la Ley General de Educación publicada el 13 de julio de 1993 y actualmente en vigencia.

En su carácter de Ley General, la actual legislación funge como una "Ley Marco", es decir, como un texto normativo que decreta la distribución de competencias que cada orden de gobierno debe ejecutar con respecto a la materia educativa. Así, la Ley General de Educación "es aquélla que puede incidir válidamente en todos los órdenes jurídicos parciales que integran al Estado mexicano" (Madero, 2015: 11) y, por lo tanto, jerárquicamente tiene influencia en las leyes federales, estatales y municipales. Esta misma Ley es la que prescribe, en su artículo 29, que el INEE es la institución que evalúa al Sistema Eductivo Nacional (SEN): "sin perjuicio de la participación que las autoridades educativas federal y locales tengan" (DOF, 19 enero 2018: 15); también la define como la autoridad en materia de evaluación educativa, coordinadora del SNEE y emisora de los lineamientos en torno a las evaluaciones. Derivado de estos mandatos, la Ley del INEE en su artículo 49 señala que: "Los lineamientos emitidos por el Instituto en materia de evaluación serán obligatorios para las Autoridades Educativas. Su incumplimiento será sancionado en términos del Título Cuarto de la

Constitución Política de los Estados Unidos Mexicanos y demás disposiciones aplicables" (INEE, 2013: 28), además de que serán nulos los procesos de evaluación realizados por las autoridades educativas en contravención a los lineamientos emitidos por el INEE.

1.2 La educación media superior en México

Para comprender el referente preciso que estas legislaciones atienden, es necesario un breve esbozo de lo que ha sido la educación media superior (EMS) en México. De esta manera, también se evidencia la estructura dentro de la que se inserta PLANEA EMS 2017.

El antecedente más directo de la EMS se remonta a 1833, año en que se estableció la Dirección General de Instrucción Pública, por iniciativa del presidente Valentín Gómez Farías y en colaboración con diversos intelectuales liberales. Con esta Dirección se abogó por la libertad de enseñanza en toda la República, adhiriéndose a la educación científica, laica y masiva. Para 1867, con la victoria definitiva de Benito Juárez, se promulgó la Ley Orgánica de Instrucción Pública. A partir de este contexto, las instituciones del nivel medio superior representaron también una respuesta a las necesidades de la sociedad mexicana. De esta manera, se fue moldeando lo que en la actualidad se conoce como bachillerato tecnológico y el bachillerato general o propedéutico.

Entre las instituciones con mayor relevancia en el devenir de la EMS en México se encuentran las siguientes:

- Escuela Nacional Preparatoria. Creada en 1867 y adscrita a partir de 1910 a la Universidad Nacional.
- Escuela Técnica de Agricultura. Fundada en 1880 y que, junto con la Escuela de Artes y Oficios, constituyen el antecedente del Bachillerato Tecnológico.
- Preparatoria Técnica. Establecida en 1931, precedente de la Enseñanza Técnica.
- Instituto Politécnico Nacional. Constituido en 1936. Fue la institución que absorbió el Departamento de Enseñanza Técnica de la Secretaría de Educación Pública (SEP).
- Preparatoria Federal por Cooperación. Promulgada en el Diario Oficial el 3 de febrero de 1940, constituyó un caso especial debido a su tipo de sostenimiento, con recursos procedentes de dependencias tanto federales, estatales, municipales, como de organizaciones sociales y particulares.
- Centro de Bachillerato Tecnológico, Agropecuario, Industrial y del Mar. Constituido en 1969.
- Colegio de Ciencias y Humanidades. Instaurado en 1971, bajo la rectoría de Pablo González Casanova.
- Colegio de Bachilleres. Creado en 1973 por decreto presidencial como organismo descentralizado del Estado.
- Consejo Nacional de Educación Técnica y Profesional (CONALEP). Fundado en 1978 bajo el enfoque de la formación y la capacitación para el trabajo.
- Determinación de la bivalencia del Bachillerato Tecnológico. Establecida en 1982.
- Centro de Estudios de Bachillerato (CEB). Bachillerato General, decretado el 23 de marzo de 1984, a partir del Plan Nacional de Desarrollo 1984-1988.
- Dirección de Educación Tecnológica Industrial (DGETI). Constituida el 16 de abril de 1971. Tomó esta denominación y se determinó que formara parte de la recién transformada Subsecretaría de Educación Media, Técnica y Superior. El 16 de enero de 2018, la DGETI se

transformó en la Unidad de Educación Media Superior Tecnológica, Industrial y de Servicios (UEMSTIS).

- Dirección General de Educación Tecnológica Agropecuaria (DGETA). Establecida el 24 de agosto de 1971 por acuerdo presidencial. El 17 de marzo de 1989 se adhirió a la DGETA la Dirección General de Ciencia y Tecnología del Mar (DGECyTM). En 1990 ambas direcciones son separadas. Finalmente, el 16 de enero de 2018, por decreto presidencial, la DGETA y la DGECyTM dieron paso a la Unidad de Educación Media Superior Tecnológica Agropecuaria y Ciencias del Mar (UEMSTAyCM).
- Sistema de Bachillerato del Gobierno de la Ciudad de México. Constituido en el año 2000.
- En el 2005, la Subsecretaría de Educación e Investigación Tecnológica (SEIT) cambia su denominación a Subsecretaría de Educación Media Superior para dar sólo atención al nivel medio superior.

Acerca de su obligatoriedad, el 9 de febrero de 2012 la educación media superior fue promulgada dentro de la educación obligatoria (EO), mediante la reforma del párrafo primero, inciso c) de la fracción II, y la fracción V del artículo 3º de la Constitución Política de los Estados Unidos Mexicanos. Esta reforma establece en sus artículos segundo y tercero transitorios que:

> **Segundo.** La obligatoriedad del Estado de garantizar la educación media superior, como deber del mismo de ofrecer un lugar para cursarla a quien teniendo la edad típica hubiera concluido la educación básica, se realizará de manera gradual y creciente a partir del ciclo escolar 2012-2013 y hasta lograr la cobertura total en sus diversas modalidades en el país a más tardar en el ciclo escolar 2021-2022, con la concurrencia presupuestal de la Federación y de las entidades federativas, y en los términos establecidos en los instrumentos del Sistema Nacional y los Sistemas Estatales de Planeación Democrática del Desarrollo.
>
> **Tercero.** Para dar cumplimiento al principio de obligatoriedad, en los presupuestos federal, de las entidades federativas y de los municipios, se incluirán los recursos necesarios; asimismo, se establecerán los mecanismos para impulsar la implementación de presupuestos plurianuales que aseguren a largo plazo los recursos económicos crecientes para infraestructura de la educación media superior (DOF, 9 febrero 2012).

Un año después, la SEP publicó el decreto de reforma en los artículos 3, 4, 9, 37, 65, 66 y adicionales al 12 y 13 de su Ley General, en los cuales estipuló la responsabilidad del Estado para brindar a los mexicanos educación obligatoria hasta el nivel medio superior. Con estas normativas, el Estado mexicano estableció su participación en el desarrollo e impartición de la EMS imponiéndose la procuración de la infraestructura necesaria para cubrir el total de la población en edad de cursarla. Desde finales del siglo pasado, debido a la amplia oferta educativa de la EMS en el país, surgió la preocupación por la unificación académica de los planes y programas de estudio y, con ello, el desarrollo de organismos capaces de coordinar la EMS en México. Por esta razón, en 1994 se creó la Dirección General del Bachillerato, la cual tenía, entre otras, la facultad de emitir la normatividad general académica en algunos subsistemas. No obstante, coordinar una estructura como la del SEN a nivel medio superior resultó un desafío, debido a sus características particulares. Este sistema abarca las modalidades de bachillerato general, tecnológico y capacitación profesional, distribuidas en 33 subsistemas con 200 planes y programas de estudio, operantes en 8 500 planteles a lo largo del país. Esta diversidad implica un reto para la movilidad de los estudiantes al interior de la EMS y para el perfil del egresado de cada subsistema.

No obstante, esta diversificación de la EMS no es única de la educación en México: la UNESCO, en la *Clasificación Internacional Normalizada de la Educación* (ISCED, por sus siglas en inglés), establece el Nivel CINE 3 (equivalente de la EMS) como la etapa en que se consolida la educación secundaria de carácter propedéutico, con el objetivo de acceder a la educación terciaria o proporcionar destrezas pertinentes para el desarrollo en el campo laboral. Por ello, generalmente, este nivel ofrece programas mayormente diferenciados, en atención a la diversidad e intereses de sus alumnos (UNESCO, 2013).

El problema, por lo tanto, no era homologar la EMS, sino tender puentes entre los subsistemas para crear identidad y un perfil de egreso más competitivo para los retos que representa el siglo XXI. Al respecto, es necesario recordar que, antes que una etapa preparatoria, la EMS es una etapa formativa de los futuros ciudadanos, porque dota a los estudiantes de un sistema de valores, de herramientas para el desarrollo psicosocial y de aptitudes para su crecimiento integral.

Por ello, con la transición a la educación por competencias y a partir de la creación de la Subsecretaría de Educación Media Superior (SEMS), el diseño de la Reforma Integral de la Educación Media Superior (RIEMS) y la instauración del Sistema Nacional de Bachillerato (SNB), se procuró materializar una dimensión pedagógica que fortaleciera el derecho de los mexicanos a recibir una enseñanza que les permita obtener los aprendizajes necesarios para enfrentar los desafíos de la actualidad.

A partir de 2008, la EMS en México se ha caracterizado por ser el espacio en el que se consolidan los aprendizajes de la educación secundaria y se prepara a los alumnos para la educación terciaria, según la *Clasificación Internacional Normalizada de la Educación,* o bien, se les provee de habilidades necesarias para el ingreso al mundo laboral. De este modo, se trata de un nivel educativo diversificado en el que se atiende la pertinencia de los aprendizajes clave, impulsando la cobertura, la calidad y la equidad de la educación. Debe recordarse que, en muchos países, incluido México, este nivel educativo marca el término de la educación obligatoria, lo que supone que los estudiantes deben tener una preparación eficiente para la continuación de estudios universitarios, el ingreso al mundo laboral y el desarrollo de habilidades socioemocionales para su desarrollo integral.

En el *Panorama Educativo en México,* 2016, se describe la estructura de la EMS:

> La educación media superior, cuya obligatoriedad se estableció en 2012 (Decreto, DOF, 9 de febrero de 2012), comprende el bachillerato general, el bachillerato tecnológico, el profesional técnico bachiller y la educación profesional técnica. La duración es variable y puede abarcar desde los dos hasta los cinco años dependiendo del modelo que se curse. En los distintos modelos educativos se prepara a los jóvenes para que posteriormente puedan cursar la educación superior, a excepción de la educación profesional técnica, que forma al estudiante para que pueda insertarse al campo laboral con sólidos conocimientos técnicos, por lo tanto, este modelo tiene un carácter terminal. No obstante, existe la posibilidad de que el alumno pueda validar materias adicionales que le permitan obtener su certificado de bachillerato. El bachillerato tecnológico forma a los jóvenes en actividades productivas agropecuarias, pesqueras, forestales, industriales, marítimas y de servicios, pero con un carácter bivalente, el cual les permite emplearse o continuar con sus estudios superiores; lo mismo sucede con el profesional técnico bachiller, que en lo sucesivo se incluye dentro del tecnológico por sus semejanzas, con la salvedad de que la preparación técnica está acotada generalmente a la industria y los servicios.

> Si se sigue la trayectoria ideal de ingreso y egreso de la educación básica, los jóvenes deberían incorporarse al primer grado de educación media superior con 15 años cumplidos al 31 de diciembre del año que inicia el ciclo escolar y concluir sus estudios a los 17 años, en los casos en los que la duración sea de tres años, que resulta la más común (INEE, 2016a: 35-36).

1.2.1 Principales organismos

El SEN "Está integrado por: educandos y educadores; autoridades educativas; planes, materiales y métodos educativos; instituciones educativas del gobierno y organismos descentralizados; instituciones de los particulares e instituciones de educación superior" (SEP, s/f). Además, aglutina la educación básica (preescolar, primaria y secundaria), la EMS (bachillerato, equivalentes y educación profesional que no requiere bachillerato) y la educación superior (estudios de técnico superior universitario, licenciatura, especialidad, maestría y doctorado) en las modalidades escolarizada, no escolarizada y mixta.

La Subsecretaría de Educación Media Superior (SEMS), organismo creado al entrar en vigor el Reglamento Interior de la SEP el 22 de enero de 2005, es responsable del "establecimiento de normas y políticas para la planeación, organización y evaluación académica y administrativa de la Educación Media Superior en sus diferentes tipos y modalidades, orientada bajo los principios de equidad y calidad, en los ámbitos Federal y Estatal" (SEMS, 2017). En su papel de garante del acceso, permanencia y calidad de la EMS, la SEMS coordinó y emprendió las acciones para iniciar el proceso de implementación de la RIEMS, que tuvo como primera etapa la publicación del Modelo de la Educación Media Superior Tecnológica (2004), por parte del Consejo Nacional para la Educación Tecnológica (COSNET), y como segunda fase la Reforma Integral de la Educación Media Superior (2008), con la cual se creó el Sistema Nacional de Bachillerato (SNB).

La RIEMS tiene como propuesta pedagógica la educación por competencias, propuesta que desde 1998 ha sido impulsada por organismos internacionales como la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y el Banco Interamericano de Desarrollo (BID), en el marco de un modelo que procura la globalización cultural y la apertura económica (Torres, 2002). Según el Acuerdo número 442, en el contexto de la RIEMS, la educación por competencias está basada en tres perspectivas: a) el planteamiento de Philippe Perrenoud, quien concibe una competencia como un trabajo mental con un objetivo práctico (la actuación adecuada y eficiente en un contexto específico), en el que intervienen recursos cognitivos; b) el enfoque de la OCDE, que destaca que el objetivo de una competencia es la resolución de demandas complejas, movilizando recursos psicosociales, y c) la perspectiva de la Asociación Nacional de Universidades e Institutos de Educación Superior (ANUIES), la cual describe a las competencias como un conjunto de conocimientos, habilidades y destrezas adquiridas de forma gradual y destinadas a satisfacer plenamente las exigencias sociales (DOF, 26 septiembre 2008). La propuesta de la RIEMS, por lo tanto, es fomentar la comprensión e integración de aprendizajes clave en la vida de los educandos, para que puedan ser aplicados en diversos contextos sociales y laborales.

De esta forma, la creación del SNB respondió a la necesidad de "Atender los problemas de cobertura, equidad y calidad que enfrentan la EMS" (SEP, 2007: 46). También surgió de la exigencia de involucrar a todos los subsistemas que conforman la EMS, para dotarlos de fundamentos teórico-prácticos

compartidos y proporcionarles una identidad común, ya que, en la medida en que cada uno de ellos posee diferentes dimensiones, estructuras y formas de organización, se hizo evidente la necesidad de establecer un perfil común al egresado de la EMS que, de acuerdo con la SEP, contribuiría para alcanzar:

> una definición universal del título de bachiller al definir las competencias que deberá poseer el egresado sin distinciones entre subsistemas [al mismo tiempo que] promueve la flexibilidad de los planes de estudio, lo cual los hace más relevantes y pertinentes (SEP, 2007: 46).

El SNB se basa en tres principios fundamentales: 1. El reconocimiento universal de todas las modalidades y subsistemas del bachillerato; 2. La pertinencia y relevancia de los planes de estudio, y 3. El tránsito entre subsistemas y escuelas (portabilidad de la educación). Asimismo, establece cuatro bases: 1. La construcción de un Marco Curricular Común (MCC); 2. La definición y reconocimiento de las porciones de la oferta de la educación media superior; 3. La profesionalización de los servicios educativos, y 4. La Certificación Nacional Complementaria.
En este contexto educativo es donde se fusionan los instrumentos de evaluación de la EMS, que desde 2008 se han desarrollado y aplicado para conocer el dominio de los aprendizajes clave que tienen los estudiantes. PLANEA EMS 2017 tiene como antecedentes a la prueba ENLACE Media Superior (ENLACE MS) y al PLANEA Media Superior 2015 y 2016 (PLANEA MS 2015-2016), antecedentes que se detallan a continuación.

1.3 ENLACE Media Superior

La prueba ENLACE MS fue una prueba nacional, objetiva y estandarizada que desarrolló el Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL) y aplicó la Dirección General de Evaluación de Políticas (DGEP) de la SEP, desde 2008 y hasta 2014, de forma anual y con una cobertura censal de la población. Su diseño se basó en el dominio restringido y la muestra de indicadores, ya que uno de sus objetivos era obtener información sobre el aprendizaje de cada estudiante y en cada escuela. Debido a la diversidad de subsistemas, ENLACE MS se enfocó en la evaluación de dos campos: Habilidad lectora y Habilidad matemática, cuyo levantamiento de datos se realizó a partir de 110 reactivos de opción múltiple, además de la aplicación de un cuestionario de contexto y, de manera matricial, una versión *pretest*.

Este instrumento de evaluación se inscribió en el Programa Sectorial de Educación 2007-2012, el cual tenía como objetivo ampliar las oportunidades educativas a partir de brindar una educación integral basada en competencias. Por ello, a partir de 2011 se ajustaron los procesos metodológicos de la prueba para alinearse a la reforma curricular de la EMS, partiendo de un cambió en el enfoque, que pasó de:

> […] un esquema orientado al desarrollo de habilidades de lectura y matemáticas —que eran comunes en los perfiles de egreso de las instituciones de educación media superior—, a uno organizado por competencias, a partir del establecimiento en 2008 del acuerdo 442 para la adopción de un Marco Curricular Común (MCC) que buscó reducir la amplia dispersión curricular de la EMS en México (Martínez, 2015: 34).

De acuerdo con el estudio *Las pruebas ENLACE y EXCALE. Un estudio de validación*, coordinado por Felipe Martínez Rizo, a partir de 2011, el marco de referencia de ENLACE MS, además del MCC de la

RIEMS, fueron los referentes de la versión anterior de la prueba y de otros, como los de PISA, TIMSS, SABER y ACREDITA-BACH, así como la documentación de la misma RIEMS. De las cuatro áreas disciplinares, ENLACE MS evaluó sólo Matemáticas y Comunicación, debido al propósito que la prueba perseguía, como se expresa en su Manual técnico:

> El diseño y propósitos de la prueba ENLACE MS se limitan a la emisión de un diagnóstico general para el sustentante, por lo que es inadecuado derivar conclusiones acerca de la EMS, los subsistemas, las escuelas o el desempeño de las entidades federativas (CENEVAL, 2013: 30).

Entre las fortalezas de ENLACE MS se deben destacar la visibilidad e interés que generó en las autoridades educativas y en los grupos preocupados por este ámbito; y la utilidad de la información de la prueba para definir programas e intervenciones de política educativa. En contraste, el estudio de Martínez Rizo y su equipo de trabajo también encontró áreas de oportunidad, sobre todo en el uso que se le dio a los resultados, alejado de la evaluación en sí misma, pues estos fueron utilizados para otorgar incentivos monetarios a los docentes y directores, y para elaborar *rankings* de escuelas (INEE, 2016b). Ambas prácticas restaron credibilidad a la prueba y terminaron por desgastarla.

1.4 Características y modalidades de PLANEA

A partir de la anterior revisión y de la necesidad de generar confianza en los resultados, surgió una nueva generación de pruebas de evaluación. Se creó el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA) que, entre otras fortalezas, tomó en cuenta las áreas de oportunidad señaladas en las pruebas precedentes y acotó la evaluación a propósitos específicos que tuvieran correspondencia con su diseño y sus rasgos técnicos, alineó los referentes con puntos centrales y estables del currículo, consideró la diversidad socioeconómica y cultural, entre otros.

PLANEA está conformado por dos modalidades en el diseño de los instrumentos y las aplicaciones, las cuales son complementarias y se aplican al finalizar cada ciclo escolar: PLANEA SEN, que aporta información nacional, estatal y por estratos escolares, y PLANEA Escuelas, que arroja información a nivel de instituciones escolares o planteles.

La primera modalidad es administrada por el INEE. Evalúa el logro referido al SEN en los distintos niveles de la EO, tiene un diseño matricial a partir de bloques de reactivos y su objetivo es conocer en qué medida los estudiantes de la EO logran un conjunto de aprendizajes clave. Brinda esta información, junto con el análisis de las brechas poblacionales, a las autoridades educativas federales, locales y organismos descentralizados para ejercer la toma de decisiones en política educativa de forma documentada. Además informa a la sociedad, tanto del logro de los estudiantes, como de las diferencias en los aprendizajes de las distintas poblaciones escolares (INEE, 2018).

La segunda modalidad, PLANEA Escuelas, es administrada por la SEP. Evalúa el nivel de los centros escolares en un solo formato de aplicación y tiene como finalidad informar a los docentes y directivos de cada centro escolar sobre el nivel de logro de aprendizajes que alcanzan sus estudiantes, lo que se espera que aprendan en los ámbitos evaluados y el tamaño del reto para la enseñanza y el aprendizaje al que se enfrentan (INEE, 2018).

1.5 PLANEA Media Superior (2015-2016)

En el ciclo escolar 2014-2015, el INEE, en coordinación con la Secretaría de Educación Pública (SEP), implementó la prueba PLANEA para la educación media superior. En 2015 y 2016 el desarrollo del instrumento lo llevó a cabo el CENEVAL y su aplicación estuvo a cargo de la SEP. Entre los rasgos que comparte con ENLACE MS se encuentran su carácter nacional, objetivo, estandarizado y censal, así como su diseño basado en el domino restringido y la muestra de indicadores. En esa ocasión se aplicaron 110 reactivos de opción múltiple, correspondientes a dos áreas disciplinares: 50 reactivos de Matemáticas y 60 de Lenguaje y Comunicación (Comprensión lectora).

A diferencia de ENLACE MS, que en el proceso de adaptación a la RIEMS mantuvo sin cambio 70% de los contenidos centrados en la medición de habilidades, PLANEA MS 2015-2016 partió de la RIEMS y del MCC para establecer los contenidos a evaluar, especificando que: "si bien se basa en los referentes del MCC, no tiene un correlato directo con contenidos curriculares o prácticas de enseñanza que pudieran estar determinadas en los planes y programas de estudio" (CENEVAL, 2016: 11).

Al mismo tiempo, la diferencia que se establece con PLANEA EMS 2017, como se verá a lo largo del presente Manual, es que la prueba de 2017 incluyó diez de las competencias disciplinares básicas en el campo formativo de Lenguaje y Comunicación, en comparación con PLANEA MS 2015 y 2016, que incluyó sólo siete de ellas. Para Matemáticas se consideraron las ocho competencias disciplinares básicas, a diferencia de PLANEA MS 2015 y 2016 con un total de seis, ya que consideró medir sólo aquellas que, con base en su criterio, no exigieran el uso de calculadoras o fórmulas especializadas. Cabe destacar que estas diferencias se deben a que PLANEA EMS 2017, además de retomar el MCC, opera con tres criterios más para determinar los contenidos a evaluar: la presencia en los planes y programas de estudio de los distintos subsistemas de la EMS; la relevancia para el dominio de los conocimientos y habilidades del campo formativo correspondiente, y la aplicación en la vida cotidiana del egresado de la EMS.

2. Conceptualización del instrumento de evaluación

La conceptualización de un instrumento de evaluación es una fase estratégica en su desarrollo, ya que es el momento en que se seleccionan los rubros centrales con respecto a su naturaleza, sus características y su alcance. En la medida en que la conceptualización se apegue a los lineamientos de la literatura especializada, se puede hablar de una contribución a la validez del instrumento de evaluación y de sus resultados. En tanto que el marco conceptual formalice la estructura del conocimiento que se pretende evaluar de manera explícita, cuidadosa y coherente con los objetivos que se pretenden alcanzar, se contribuye a una mayor equidad en la medición.

La prueba PLANEA EMS 2017 siguió un diseño apegado a los criterios para el desarrollo de instrumentos de evaluación establecidos por el INEE, cuya primera versión se publicó en 2014 con el título de *Criterios técnicos para el desarrollo y uso de instrumentos de evaluación*, y fue elaborada considerando los estándares de calidad técnica propuestos por organismos de evaluación educativa internacionales y nacionales, como la European Commission (EC); el United Nations Evaluation Group (UNEG); el Committee for the Development of Educational and Psychological Testing Standards de la American Educational Reasearch Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME); el Joint Committee on Testing Practices (JCTP), el Educational Testing Service (ETS), el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) y el Instituto Nacional para la Evaluación para la Educación (INEE, 2014). En 2017 se reformuló el documento con el nombre de *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*.

El propósito de estos criterios es proporcionar un marco de referencia para garantizar que durante el desarrollo de la prueba se consideren e incluyan los aspectos determinantes de la evaluación y que exista un punto de partida en cuanto a los fundamentos teóricos y sus orientaciones que doten de validez los resultados de las evaluaciones arrojados por los instrumentos de evaluación.

2.1. Planeación general del instrumento de evaluación

Es el primer paso que señala este documento guía. Aquí se definen las características de la evaluación, tomando como eje orientador el marco de referencia en el que se inserta esta nueva generación de instrumentos de evaluación, principalmente los planes y programas de estudio y el Marco Curricular Común (MCC), su representatividad y la relevancia de los contenidos que se pretende evaluar, ya que a partir del análisis formal de los contenidos curriculares (definición del universo de contenido), se pueden determinar aquellos que se considerarán para evaluar (la definición del universo de medida). Así pues, en esta etapa es importante mostrar el proceso que siguieron los responsables del diseño y construcción de la prueba para determinar los contenidos de la evaluación en 2017.

Además del currículo, se exponen los propósitos de PLANEA EMS 2017, la población objetivo a la cual va dirigida la prueba, las condiciones de factibilidad y el uso de resultados. También se perfilan los cuerpos colegiados, indispensables en la construcción del instrumento de evaluación, ya que su formación, su experiencia profesional y su función al interior de la prueba favorecen el desarrollo de un instrumento cuyos resultados permiten hacer inferencias válidas.

2.1.1. Propósito de PLANEA EMS

Como ya se mencionó, al diseñar un instrumento de evaluación, la delimitación de los objetivos que guían su construcción adquiere relevancia, porque el diseño de la prueba debe ser claro, acotado y congruente con éstos. Por ello, resulta importante visualizar estos objetivos que se mencionan en el documento rector de PLANEA, mismo que establece la finalidad de estas evaluaciones, la cual se cumple de manera óptima si: "los objetivos específicos son definidos con claridad; los usuarios de la información conocen sus alcances; [...] y se verifica que la información que ofrecen sus resultados corresponde con los propósitos de su diseño" (INEE, 2018: 9).

Por ello, el propósito general de PLANEA EMS 2017 es evaluar el nivel medio superior del Sistema Educativo Nacional (SEN) para conocer en qué medida los estudiantes de este grado terminal (bachillerato general, bachillerato tecnológico y educación profesional técnica) adquieren los aprendizajes clave en las áreas disciplinares de Lenguaje y Comunicación y Matemáticas, ya que:

> [...] integran herramientas esenciales para la vida diaria y para el desarrollo de aprendizajes en otras áreas del conocimiento. Asimismo, estas dos áreas representan una buena proporción de la carga curricular en educación básica y media superior. Por lo anterior, los aprendizajes en estos dos campos son un buen indicador de los resultados educativos en general (INEE, 2018: 15).

Estas áreas se evalúan a partir de cuatro ejes temáticos. En el caso de Lenguaje y Comunicación: Manejo y construcción de la información, Texto argumentativo, Texto expositivo y Texto literario, porque favorecen el desarrollo de competencias tanto de producción como de comprensión de textos, que inciden directamente en el desarrollo de la capacidad comunicativa de los alumnos. En el área de Matemáticas son: Sentido numérico y pensamiento algebraico; Cambios y relaciones; Forma, espacio y medida y Manejo de la información, porque agrupan conocimientos que exploran la capacidad de los estudiantes para emplear y transformar sus aprendizajes en herramientas para interpretar, comprender, aplicar, analizar, sintetizar y evaluar matemáticamente su entorno, así como dar solución a diferentes problemas mediante el razonamiento.

PLANEA EMS 2017 también explora habilidades socioemocionales que "se relaciona[n] con la planificación, la toma de decisiones y la solución de problemas de distinta naturaleza" (INEE, 2018 18). De estas habilidades, la prueba explora cuatro escalas: Persistencia, Empatía, Manejo del estrés y Toma de decisiones. Además, recaba información sobre las condiciones escolares, familiares y personales en que se encuentran los alumnos para poder hacer una devolución contextualizada de resultados, todo ello a través de cuestionarios de contexto que se aplican a directivos, docentes y estudiantes.

2.1.2. Población objetivo

En términos generales, la prueba PLANEA se aplica en los grados terminales de cada nivel escolar a lo largo de la educación obligatoria (EO), debido a que esto constituye un indicador de la eficiencia del proceso educativo en su conjunto. Los aprendizajes clave que se evalúan son acumulativos, lo que permite reconocer los logros de los alumnos a lo largo de varios años, durante los cuales éstos van integrando diferentes aprendizajes que "conforman una red compleja de conocimientos, habilidades y competencias" (INEE, 2015b: 15). Esto adquiere relevancia en la medida en que

también apoya la sistematización de estrategias encaminadas a alcanzar el máximo logro de aprendizajes por parte de los estudiantes.

La población objetivo de PLANEA EMS 2017 está constituida por estudiantes del último grado de la educación media superior (EMS) de los diferentes subsistemas en todo el país, instituciones centralizadas del gobierno federal, descentralizadas de la federación, descentralizadas de las entidades federativas, estatales, de carácter autónomo y planteles particulares con validez oficial de la Secretaría de Educación Pública (SEP).

Tabla 1. Subsistemas de educación media superior que conforman la población objetivo de PLANEA EMS 2017

Sostenimiento	Control administrativo	Instituciones	Modelo educativo
Federal	Centralizado (SEMS)	UEMSTIS (antes DGETI)	CETIS, CBTIS
		DGETAS	CBTA, CBTF
		DGECyTM	CETMAR, CETAC
		DGB	CEB, PREFECOS, Preparatoria federal "Lázaro Cárdenas"
	Descentralizados	CONALEP (CDMX y Oaxaca) CETI Guadalajara COBACH (CDMX)	Profesional técnico Bachiller Bachillerato tecnológico Bachillerato general
	Desconcentrados	INBA	Bachillerato general
		IPN	Bachillerato tecnológico
Estatal	Centralizados	TELEBACH, Telebachilleratos, comunitarios, Bachilleratos estatales	Bachillerato general
	Descentralizados	CECYTE	Bachillerato tecnológico
		COBACH	Bachillerato general
		CONALEP	Profesional técnico o Profesional técnico Bachiller
Autónomos		Escuelas asociadas con las Universidades Públicas Autónomas	Bachillerato general y tecnológico
Privados	Subsidiados	Organismos subsidiados por los estados, SEP y Asociaciones Civiles (PREECOS y PREFECOS)	Bachillerato general y tecnológico
	No subsidiados	Bachilleratos particulares y centros de estudios tecnológicos particulares	Bachillerato general y tecnológico Bachilleratos incorporados

Fuente: INEE (2011). *La Educación Media Superior en México. INFORME 2010-2011*.

2.1.3. Condiciones de factibilidad de la prueba

Las condiciones de factibilidad aluden a los recursos necesarios y disponibles para alcanzar los objetivos propuestos. Si se observa que las evaluaciones de logro educativo impactan directamente en el derecho de los ciudadanos a recibir una educación de calidad, así como en las obligaciones del SEN para contribuir a la formulación de políticas educativas y de fomentar la transparencia y la rendición de cuentas, entonces es importante establecer las condiciones de factibilidad a partir de las cuales se planifican los instrumentos de evaluación, tanto en los aspectos técnicos como en los organizacionales, ya que en la medida en que éstos tengan un mayor grado de viabilidad, mayor será la garantía para lograr los objetivos que se han propuesto.

En el caso de PLANEA EMS 2017, detrás de su planeación existe una legislación que establece las pautas que deben seguir las evaluaciones con este alcance, al mismo tiempo existe una literatura especializada que otorga los lineamientos técnicos y los estándares a seguir en el proceso metodológico. Por ende, se trata de un instrumento que, como se mencionó anteriormente, se basa en los criterios para el desarrollo de instrumentos establecidos por el INEE desde 2014. Esto promueve la calidad técnica de la prueba, ya que cuenta con un marco de referencia que funciona como soporte para la planeación de su desarrollo.

Además de lo anterior, los antecedentes como ENLACE Media Superior y PLANEA MS 2015-2016 permitieron retomar áreas de oportunidad para garantizar el rigor metodológico del diseño, la construcción y la aplicación de la prueba, así como fortalecer la correspondencia entre la información que ofrecen los resultados y los propósitos que se plantearon durante el diseño para evitar el sesgo que pudiera resultar de la diversidad genérica, étnica, cultural y lingüística del país.

Dentro de estas condiciones de factibilidad para el diseño y construcción del instrumento de evaluación, también debe ser considerado el personal que tuvo a su cargo la ejecución de la prueba, pues su trabajo estuvo enfocado en la consecución de los objetivos propuestos. Los profesionistas que colaboraron en el desarrollo de la prueba cuentan con la formación académica idónea y amplia experiencia en el nivel medio superior como expertos en la enseñanza de las asignaturas Lenguaje y Comunicación y Matemáticas, como formadores de profesores de bachillerato, coordinadores de grupos para la elaboración de material educativo, investigadores o líderes de proyectos educativos con experiencia en diversas áreas del desarrollo de instrumentos de evaluación.

2.1.4. Perfiles de los integrantes de los consejos y comités

En el diseño de PLANEA EMS 2017 participaron, además de los coordinadores académicos (los jefes de prueba y la subdirectora adscritos a la Dirección de Evaluaciones Nacionales de Resultados Educativos, del INEE), seis comités más que colaboraron en diversos momentos de la construcción de las pruebas de Lenguaje y Comunicación o Matemáticas. Los integrantes de estos grupos son expertos dentro del campo a evaluar.

A continuación se mencionan estos comités, los cuales se detallarán en los apartados dedicados a ellos.

1. Comité Académico
2. Comité de Especificaciones
3. Comité de Elaboración de Reactivos
4. Comité de Validez y Sesgo
5. Comité de Niveles de Logro y Puntos de Corte

Debido al carácter teórico y curricular del trabajo que realizó el Comité Académico, el perfil profesional incluyó alguna de las siguientes competencias:

- Ser académico especialista en el campo disciplinar de Lenguaje y Comunicación o Matemáticas.
- Ser investigador del nivel medio superior con acreditado sustento mediante publicaciones o trabajos referentes a este nivel escolar.

- Ser docente en funciones, conocedor de los planes y programas de estudio del subsistema representado.
- Ser conocedor del Marco Curricular Común desde la docencia, la investigación o la administración.

Para los siguientes tres comités (Especificaciones, Elaboración de Reactivos y Validez y Sesgo), el perfil profesional solicitado comprendió alguna de las siguientes características:

- Ser profesional experto en el Marco Curricular Común de las áreas disciplinares de Lenguaje y Comunicación o Matemáticas.
- Ser especialista en didáctica con amplia experiencia docente en EMS o con experiencia en la formación de docentes.
- Ser experto en el enfoque por competencias.
- Ser autor de libros sobre temas del bachillerato.
- Ser especialista del enfoque intercultural.
- Ser conocedor del nivel medio superior desde la docencia, la investigación o la administración.

El Comité de Niveles de Logro y Puntos de Corte tuvo un trabajo dividido en dos grupos (A y B), para los cuales se solicitaron dos perfiles distintos. Para el primer grupo, que redactó los descriptores de los niveles de logro, se solicitó que los expertos tuvieran las siguientes aptitudes:

- Ser especialista conocedor del currículo de la EMS, desde el MCC o de los planes y programas de estudio de los diferentes subsistemas.
- Ser investigador educativo del nivel medio superior.
- Tener amplia experiencia en la elaboración de informes de resultados.

Para el Grupo B, que estableció los puntos de corte, el perfil solicitado abarcó alguna de las siguientes competencias:

- Ser docente frente a grupo del nivel medio superior.
- Ser autor de libros sobre temas del bachillerato.
- Ser especialista en el campo disciplinar de Lenguaje y Comunicación o Matemáticas.

En la conformación de estos comités se observó una cuidadosa atención para que fuera representativa de los diferentes subsistemas de la EMS, atendiendo a la multiculturalidad del país y para evitar problemas de sesgo desde la etapa más temprana.

En el Anexo I se puede consultar a detalle la tabla de perfiles profesionales y académicos de los integrantes en los consejos y comités que elaboraron el instrumento de evaluación.

2.1.5. Usos previstos

La prueba PLANEA EMS 2017 proporciona información respecto al nivel medio superior del SEN y muestra las particularidades entre las poblaciones de estudiantes de los diferentes subsistemas que lo conforman. Sus resultados pueden contribuir a la toma de decisiones para mejorar la calidad de

la educación y para que los estudiantes logren integrar los aprendizajes clave al concluir el último ciclo escolar de su EO. Los destinatarios de la información son:

a) Las autoridades educativas, que tienen como función la toma de decisiones, monitoreo, planeación, programación y operación del sistema educativo y de los centros escolares.
b) Las comisiones parlamentarias relacionadas con el ámbito educativo, con la finalidad de que puedan conocer y valorar los avances y retrocesos en el logro de los aprendizajes a nivel nacional y apoyar, desde el ámbito legislativo, la mejora educativa mediante el diseño de políticas públicas fundamentadas.
c) Los académicos e investigadores, quienes asumen la tarea de contribuir al desarrollo de innovaciones para la mejora de las pruebas y explorar estrategias para fortalecer las prácticas de evaluación en el aula.
d) Los docentes y personal de apoyo técnico pedagógico, con el fin de que logren identificar las fortalezas y debilidades de los aspectos curriculares e implementar estrategias para enfocar las prácticas docentes en la mejora del aprendizaje.
e) Los estudiantes, actores principales, cuyo involucramiento activo es un aspecto clave para el desarrollo de sus propios aprendizajes.
f) Las familias, que tienen el derecho de conocer y comprender las evaluaciones, para apoyar a sus hijos e hijas en el proceso de enseñanza-aprendizaje.
g) Los medios de comunicación, que apoyan con la difusión de resultados y con el desarrollo de capacidades de análisis e interpretación de las evaluaciones en la ciudadanía.
h) Las organizaciones civiles y la sociedad en general, para tomar decisiones informadas a fin de hacer exigible el derecho que tienen las niñas, niños y jóvenes a una educación de calidad (INEE, 2016b).

2.2. Diseño del instrumento de evaluación

La fase de diseño de la prueba PLANEA EMS 2017 constituyó tanto la definición conceptual como operacional de los atributos medibles en el instrumento de evaluación (el objeto de medida). Se determinó y justificó la relevancia, pertinencia y representatividad de los aprendizajes que conformaron la prueba, así como la metodología y los criterios que guiaron la construcción de las especificaciones base para la posterior elaboración de los reactivos o ítems.

Seguir puntualmente los criterios establecidos por el INEE resultó fundamental para PLANEA EMS 2017, ya que a partir de los mismos se sustentaron la fortaleza, la confiabilidad y la validez de la prueba; ésta, a su vez, arrojó resultados e informes nacionales y estatales útiles para la rendición de cuentas y la toma de decisiones en política educativa.

2.2.1. Marco de referencia para el desarrollo del instrumento

PLANEA EMS 2017 se enmarca en el modelo instaurado desde la RIEMS, el cual estableció cuatro desempeños terminales: a) competencias genéricas, b) competencias disciplinares básicas, c) competencias disciplinares extendidas, de carácter propedéutico, y d) competencias profesionales para el trabajo (DOF, 26 septiembre 2008). Las dos primeras, de acuerdo con el documento rector de la RIEMS, tienen como función darle unidad a la oferta de la EMS pues son comunes a todos los subsistemas; las dos últimas se definen de acuerdo con los objetivos y necesidades de cada institución, particularizando a cada subsistema.

Esta visión de la educación en México se expresa a través de dos características, que en este caso definen a las competencias genéricas: que sean conocimientos clave (relevantes a lo largo de la vida) y que puedan ser aplicados en contextos tanto personales y sociales como en académicos y laborales. Además, la educación también debe centrarse en aprendizajes transversales, relevantes a todas las disciplinas académicas, a los procesos escolares de apoyo y a las actividades extracurriculares de los estudiantes, y el conocimiento debe ser transferible, para facilitar y reforzar la adquisición de otras competencias (DOF, 26 septiembre, 2018).

Así, el desglose de la propuesta pedagógica de la RIEMS, concentrada en el Acuerdo número 442, es uno de los referentes para la construcción de PLANEA EMS 2017, pues sienta las bases para establecer cómo se conceptualiza el aprendizaje significativo y las habilidades reflexivas, de solución de problemas y de toma de decisiones, a partir de la pertinencia de los aprendizajes de la EMS, estrechamente relacionada con las exigencias de la actualidad y del medio en el que se lleva a cabo el proceso de aprendizaje: un mundo cada vez más global, empresarial, tecnológico y que exige, por tanto, un perfil dirigido al desarrollo competente dentro de este contexto.

Al mismo tiempo, PLANEA EMS 2017 tuvo como referente el MCC y los planes y programas de estudio de los distintos subsistemas de la EMS (debidamente representados por los expertos que conformaron el Comité Académico). La decisión de considerar tanto las competencias señaladas en el MCC como el contenido de los planes y programas se originó en algunos estudios realizados previamente a instrumentos de evaluación anteriores, en los que se analizó la dificultad que, históricamente y en diversos niveles educativos, ha significado el currículo en México: "la falta de estabilidad del currículo es una desventaja para cualquier prueba referida a un criterio orientado por dicho currículo" (Martínez, 2015: 34).

La prueba PLANEA EMS 2017 contempló el MCC y los planes y programas de estudio bajo el entendido de que existe un cruce entre ambos referentes que no puede ser soslayado, ya que la instrucción que reciben los alumnos, aunque está guiada por el MCC en el contexto de la RIEMS, tiene como soporte el plan de estudios propio de cada subsistema.

Por otro lado, la importancia de tomar las experiencias de los instrumentos de evaluación que operaron anteriormente al PLANEA EMS 2017 hizo necesario también tomar como referente la documentación de EXCALE, ENLACE MS y PLANEA MS (2015-2016). La justificación para retomar la experiencia de EXCALE se debe a que su objetivo general, que era informar sobre el nivel de aprendizaje en el sistema educativo e identificar los factores que impactan en él para apoyar las decisiones de política educativa, se relaciona con el objetivo de PLANEA EMS 2017 (guardando las debidas distancias). Además, ambas pruebas comparten algunas características que resultan definitorias para su construcción, por ejemplo, están alineadas al currículo (en el caso de PLANEA EMS 2017, éste se alineó a contenidos comunes que se consideraron relevantes para la disciplina), son criteriales y muestrales, y cuentan con cuestionarios de contexto. Mientras que las de las dos últimas pruebas resultan importantes ya que fueron éstas las que transitaron de la versión del dominio de habilidades básicas al de la evaluación de competencias disciplinares. En este sentido, tanto ENLACE MS (a partir de 2011) como PLANEA MS 2015-2016 buscaban contribuir a la transición hacia la educación por competencias.

Por otra parte, en la etapa operacional tanto de la construcción como de la validación de reactivos, PLANEA EMS 2017 se alineó a los *Manuales técnicos* establecidos por el INEE, los cuales fueron construidos con base en experiencias de organismos especializados en el área de la evaluación para

la educación, cuyo origen se remonta a 1954, con el Comité Conjunto sobre Prácticas Evaluativas. Así, por ejemplo, en el *Manual técnico de Validación de reactivos* (INEE, 2006) se dice que, así como en otras partes del mundo se han publicado versiones de estándares de calidad para las evaluaciones educativas, el INEE también contribuía con esta tarea al presentar sus criterios técnicos para normar la construcción, evaluación y documentación en materia de evaluación. En este sentido, apegarse a dichos criterios apoya la validez de los resultados e interpretaciones arrojadas por la prueba.

El documento rector de PLANEA también fue clave en el diseño de la prueba a lo largo de sus diferentes fases. Este documento, surgido de la necesidad de ser, precisamente, el marco general de referencia, orientación y presentación de PLANEA y de sus diversas modalidades, agrupa los fundamentos, los antecedentes, los propósitos, las modalidades, los aprendizajes que se evalúan, y las estrategias de la difusión y uso de resultados, por lo que resulta obligatorio su seguimiento, en virtud de que, aunque las distintas modalidades de PLANEA se orientan a la obtención de informaciones específicas (ver el inciso 3.4 de este Manual), existe una integración lógica y articulada de éstas. Por esta razón, es necesario guiarse con los documentos rectores que sintetizan el perfil y los objetivos del Plan Nacional para la Evaluación de los Aprendizajes.

2.2.2. Delimitación conceptual del objeto de medida y contenidos a evaluar

Una vez establecido el marco de referencia que guió el instrumento de evaluación, comenzó la primera tarea encaminada a concretar los aprendizajes a evaluar. En este sentido, debe entenderse que, de acuerdo con los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, el objeto de medida se define como el "conjunto de características o atributos que se miden en el instrumento de evaluación" (INEE, 2017: 8). En el caso de PLANEA EMS 2017, el propósito fue evaluar el dominio de aprendizajes clave de los alumnos de los grados terminales de la EMS, en sus diferentes subsistemas y a partir de dos áreas disciplinares: Lenguaje y Comunicación y Matemáticas, ya que los contenidos que se trabajan en éstas resultan esenciales para la continua adquisición de nuevos aprendizajes, son relevantes para el dominio del campo disciplinar y prevalecen en el tiempo a pesar de los cambios realizados a estos últimos.

Delimitar el objeto de medida de la prueba y establecer los indicadores de aprendizaje (que posteriormente dieron lugar a los ítems que definieron la estructura de la prueba) planteó importantes retos, debido a la heterogeneidad de la EMS. Por esta razón fue necesario contemplar tanto los diferentes referentes curriculares (los planes y programas de estudio de los distintos subsistemas) como el perfil de egreso establecido en el MCC. Al mismo tiempo, hubo que centrarse en las características de los aprendizajes clave, enmarcados en el enfoque de la educación por competencias. Para solventar estos requerimientos, PLANEA EMS 2017 se valió de cuatro criterios para determinar los contenidos y aprendizajes clave a incluir en el instrumento de evaluación; dichos criterios procuraron integrar las exigencias y los objetivos particulares de la prueba:

1. Correspondencia con las competencias disciplinares básicas de Lenguaje y Comunicación y Matemáticas, y su vinculación con las competencias genéricas.
2. Presencia en los planes y programas de estudio de los distintos subsistemas de la EMS.
3. Relevancia para el dominio de los conocimientos y habilidades del campo formativo correspondiente.
4. Aplicación en la vida cotidiana del egresado de EMS (sentido práctico del aprendizaje).

Los contenidos que cumplieron con estos cuatro criterios fueron los que integraron la Tabla de contenidos y se operacionalizaron en las especificaciones, ya que demostraron ser contenidos que fomentan el aprendizaje significativo, que poseen un objetivo práctico y que, en general, movilizan los recursos de los estudiantes para enfrentar las exigencias de su entorno.

2.2.3. Ficha técnica de la prueba

PLANEA EMS 2017 es un instrumento de evaluación diseñado para la modalidad de PLANEA SEN, la cual se centra en proporcionar información al SEN, a nivel nacional, estatal y por estratos escolares. Se trata de una prueba matricial y muestral referida a currículo, cuyos referentes, como se expuso con anterioridad, son el MCC y la presencia de estos contenidos en los planes y programas de los distintos subsistemas de la EMS. La prueba incluye varios formatos de aplicación, distribuidos en cuadernillos de 50 reactivos para cada área disciplinar (Lenguaje y Comunicación y Matemáticas) que se aplican a una muestra representativa de la población objetivo, esto debido a que un diseño matricial, a diferencia de uno censal, apoya la evaluación de un mayor número de contenidos curriculares, con lo que se logra una mayor precisión en la medición.

La ficha técnica que a continuación se presenta, y que funge como evidencia para la verificación documental, fue resguardada por los coordinadores académicos de la prueba, y presenta de manera esquemática las especificaciones técnicas del instrumento de evaluación, de acuerdo con los señalamientos que apuntan los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, del INEE.

Ficha técnica de PLANEA EMS 2017	
Tipo de instrumento	Diagnóstico, de bajo impacto
Dominio explorado	Competencias de las áreas disciplinares de Lenguaje y Comunicación y Matemáticas
Modelo de diseño	Matricial
Longitud del instrumento	130 indicadores de aprendizaje para Lenguaje y Comunicación y 102 para Matemáticas
Características específicas del instrumento	Aplicación de cuadernillos con reactivos de opción múltiple en lápiz y papel con un alto control de aplicación
Modalidad	PLANEA SEN
Propósito	Conocer la medida en que los estudiantes logran el dominio de un conjunto de aprendizajes clave en la EMS
Cobertura geográfica	Nacional
Periodicidad	Trianual
Cobertura de la población	Muestral
Población objetivo	Estudiantes de grados terminales del nivel medio superior.
Uso y alcance de sus resultados	Apoyar la toma de decisiones, monitoreo, planeación, programación y operación del sistema educativo y de los centros escolares, desde el ámbito de la docencia, la investigación y la legislación
Tratamiento que se le dará a las respuestas	Análisis psicométrico Análisis de comportamiento diferencial (DIF)
Marco teórico de la medición	Teoría clásica del test Modelo logístico de Rasch-TRI (Teoría de respuesta al Ítem)
Puntuación del instrumento	Cuatro Niveles de Logro (I, II, III y IV)
Usuarios de la información	Autoridades educativas, comisiones parlamentarias, académicos, investigadores, docentes, personal de apoyo técnico pedagógico, estudiantes, familias mexicanas, medios de comunicación, organizaciones civiles y sociedad en genera

2.3. Comité Académico de Determinación de Contenidos e indicadores de aprendizaje de PLANEA EMS 2017

El Comité Académico de Determinación de Contenidos fue el grupo colegiado responsable de establecer tanto el objeto de medida que integraría la prueba PLANEA EMS 2017, como la descripción de los indicadores de aprendizaje que se evaluaron, los cuales debieron corresponder con los objetivos específicos de la prueba. Por ello, metodológicamente y como lo establecen los *Criterios técnicos*, fue el primer grupo de expertos que sesionó ya que, a partir de su trabajo conceptual, PLANEA EMS 2017 pudo ser operacionalizado en los subsecuentes comités.

Debido a que el trabajo de este Comité tiene una fundamentación curricular, fue importante que estuviera conformado por profesores especialistas en la enseñanza del campo disciplinar (Lenguaje y Comunicación o Matemáticas), por investigadores del nivel medio superior, por docentes en funciones y por expertos en el MCC; del mismo modo, se procuró que las 32 entidades federativas tuvieran representación, así como los diferentes subsistemas, con la finalidad de tener una amplia integración de la EMS a nivel nacional.

Durante la conformación de este Comité se puso especial énfasis en la diversidad cultural y lingüística del país, ya que en un contexto multicultural como el que representa los Estados Unidos Mexicanos, se debe promover la educación intercultural para evitar posibles fuentes de sesgo desde la propia conceptualización del instrumento de evaluación. Por esta razón, y atendiendo a las recomendaciones del equipo de trabajo de Martínez Rizo (2015), la prueba PLANEA EMS 2017 contó con la colaboración de profesores adscritos a diversas instituciones de educación media superior, los cuales colaboraron con la integración de los contenidos que resultaban significativos desde su experiencia en estas áreas.

En la tabla 2 se enlistan las instituciones que estuvieron representadas en el Comité Académico de Determinación de Contenidos e indicadores de aprendizaje.

Tabla 2. Instituciones de educación media superior representadas en el Comité Académico de PLANEA EMS 2017

Dirección General de Bachillerato	Centro de Estudios Científicos y Tecnológicos
Dirección General de Educación Tecnológica Industria	Telebachillerato Comunitario
Dirección General de Educación Tecnológica Agropecuari	Instituto de Educación Media Superior
Dirección General de Centros de Formación para el Trabajo	Bachillerato del Estado de México, Universidad Autónoma del Estado de México
Dirección General de Educación en Ciencia y Tecnología del Mar	Escuela Nacional Preparatoria (UNAM)
Colegio de Bachilleres **CDMX**	Colegio de Ciencias y Humanidades (UNAM)
Colegio de Bachilleres Tabasco (Bachillerato intercultural)	Consejo Académico del Bachillerato (UNAM)
Colegio Nacional de Educación Profesional Técnica **CDMX**	Universidad Pedagógica Nacional
Colegio Nacional de Educación Profesional Técnica estatal	Sociedad Matemática Mexicana

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

2.3.1. Procedimiento metodológico para el desarrollo del Comité Académico

El procedimiento metodológico del Comité Académico se constituyó por tres fases: la primera de ellas consistió en consignar en una matriz o tabla de contenidos, las categorías temáticas a medir y definir los elementos de la estructura de la prueba a partir de un análisis formal que permitiera determinar el universo de medida, acorde con los objetivos de la prueba. Fue fundamental que la construcción de la Tabla de contenidos se apegara a los referentes curriculares que estipularon los conocimientos esperados de la población objetivo, para asegurar que se evaluaran contenidos que permitieran verificar el dominio de conocimientos y habilidades de la población que fue producto del proceso de enseñanza-aprendizaje. Se tuvo cuidado de no evaluar contenidos que no fueran abordados en el mismo, ya que esto podía comprometer la validez de los resultados de la prueba.

Los integrantes del Comité analizaron, a partir de estos criterios, las competencias disciplinares básicas que contribuían a la consolidación de los atributos de las competencias genéricas (anexo II). Con este fin se elaboró una matriz de doble entrada que orientó a las competencias disciplinares de manera vertical, en columnas, y a cada competencia genérica con sus atributos de manera horizontal, por filas, como se muestra en el formato 1.

Formato 1. Competencias disciplinares básicas que apoyan la consolidación de los atributos de las competencias genéricas

Competencias genéricas / Competencias disciplinares	Perfil A. Se autodetermina y cuida de sí.					
	CG1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.					
CDM8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.	1. Enfrenta las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.	2. Identifica sus emociones, las maneja de manera constructiva y reconoce la necesidad de solicitar apoyo ante una situación que lo rebase.	3. Elige alternativas y cursos de acción con base en criterios sustentados y en el marco de un proyecto de vida.	4. Analiza críticamente los factores que influyen en su toma de decisiones.	5. Asume las consecuencias de sus comportamientos y decisiones.	6. Administra los recursos disponibles teniendo en cuenta las restricciones para el logro de sus metas.
Consolidación de los atributos						
Contenido Temático						

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Una vez hecha esta disposición, los integrantes del cuerpo colegiado anotaron en la celda producto del cruce su respuesta a la siguiente pregunta:

¿La competencia disciplinar *X* permite consolidar el atributo *Y* de la competencia genérica *Z*?

Sí = 1 No = 0

Una vez realizada esta operación, el Comité Académico pudo determinar los contenidos específicos (temas, conceptos, procedimientos) de cada competencia disciplinar que permitía consolidar los atributos correspondientes, de acuerdo con las celdas marcadas en la matriz anterior. Con esto como base, el cuerpo colegiado realizó, desde su conocimiento de los planes y programas de estudio

de cada subsistema, un análisis formal acerca de estos contenidos específicos, para determinar si eran comunes a los diferentes subsistemas. Así se evitó que existieran estudiantes en desventaja con respecto a la evaluación. En esta misma fase, y una vez realizado este cruce del MCC con los planes y programas, se analizó la pertinencia de incluir contenidos que fueran importantes desde el punto de vista disciplinar y del aprendizaje de los estudiantes, ya porque formaran parte de las competencias disciplinares o de los planes y programas, o porque por alguna razón no hubieran sido ubicados en la matriz elaborada durante la primera fase.

Finalmente, en una tercera etapa, el Comité Académico describió los indicadores de los aprendizajes clave a evaluar que, en términos operativos, se traducen como la unidad básica del objeto de la evaluación y que tienen como meta operacionalizar los objetivos a evaluar en componentes sintéticos que aporten evidencia del aprendizaje de los estudiantes. PLANEA, en su Documento rector, considera que los aprendizajes clave:

a) Son fundamentales para el dominio de los conocimientos y habilidades del campo de formación correspondiente.
b) Prevalecen en el currículo, más allá de las modificaciones que éste suele sufrir a lo largo del tiempo.

Los aprendizajes clave, además, son útiles en la vida práctica y ciudadana, aunque existen algunos que, por su nivel de abstracción, no cumplen con esta condición, pero resultan indispensables para seguir avanzando en el campo de formación correspondiente (INEE, 2018: 15).

- **Criterios para la descripción de indicadores de aprendizajes clave a evaluar:**

Al ser elementos que informan sobre el estado de la EMS en México, los indicadores de aprendizaje, por ende, deben desagregar analíticamente los contenidos que previamente determinó el Comité Académico. Para realizar este procedimiento es importante que se considere el propósito de la prueba, la población objetivo y, en general, el contexto social, cognitivo y afectivo en el que se inscribe el proceso de enseñanza-aprendizaje.

Uno de los criterios que fue considerado al momento de describir los indicadores de aprendizajes clave fue su pertinencia con respecto al conjunto mayor de los contenidos específicos, dicha relevancia se observó a partir de determinar si el contenido era requisito para la adquisición de otros, si en éste influían otros contenidos que deberían ser integrados por él o si tenía una relevancia dentro de la disciplina (Lenguaje y Comunicación o Matemáticas).

También se cuidó que la redacción estuviese libre de sesgo lingüístico y cultural, que fuera concreta y que expresara de forma directa el aprendizaje o desempeño a evaluar en los contenidos, a partir de verbos que indicaran la acción concreta que el estudiante debía realizar para contestar correctamente el contenido curricular.

Como resultado de la primera fase de análisis, el Comité Académico de Lenguaje y Comunicación determinó que sólo se elaborarían indicadores de aprendizaje para 10 de las 12 competencias disciplinares básicas establecidas en el MCC. La decisión de no incluir a dos de ellas obedeció a que una está relacionada con el aprendizaje de una segunda lengua y la otra con el uso de las tecnologías de la comunicación. Así, los contenidos a evaluar se organizaron alrededor de cuatro ejes temáticos: Manejo y construcción de la información, Texto argumentativo, Texto expositivo y Texto literario.

Para el caso de Matemáticas, el Comité propuso indicadores de aprendizaje para las 8 competencias disciplinares básicas establecida en el MCC y los contenidos se organizaron en los ejes temáticos: Sentido numérico y pensamiento algebraico, Cambios y relaciones, Forma, espacio y medida, y Manejo de la información.

- **Descripción y revisión de indicadores de aprendizajes clave**

Para llevar a cabo la descripción de los indicadores de aprendizaje se presentaron ante el Comité los materiales de apoyo (tabla de contenidos y ejemplos de la redacción de indicadores de aprendizaje clave) y se procedió a exponer el procedimiento para realizar la tarea, la cual inició con la revisión de los contenidos propuestos para cada competencia disciplinar, por tratarse del núcleo del que partían los indicadores de aprendizaje y al cual también representaban.

Para determinar el indicador de aprendizaje, los expertos de cada área dieron respuesta a la siguiente pregunta: ¿Qué debe aprender el alumno sobre el tema *X* para que esto mismo que aprende lo ayude a desarrollar el atributo *Y*? A partir de la respuesta de los integrantes del Comité, se precisó la acción (¿Qué hace el estudiante?) que éste debió llevar a cabo para dar muestra de que era capaz de responder correctamente a un contenido específico (¿Qué es lo que hace en concreto?), como se evidencia en el siguiente indicador ejemplo, que se desagrega del contenido "Planificación de un escrito":

Organiza / la redacción de un texto de acuerdo a su propósito comunicativo.
(acción) (contenido)

Como se observa, el desglose analítico consistió en determinar que dentro de la Planificación de un escrito era necesario que el estudiante pudiera organizar la redacción de un texto de acuerdo con su propósito comunicativo, ya que este aprendizaje contribuía a que el alumno aprendiera a "Administrar los recursos disponibles teniendo en cuenta las restricciones para el logro de sus metas" (Atributo 6 de la Competencia genérica 1).

Es importante tener en cuenta que en este ejercicio también se consideró el contexto en el que se manifestaría el indicador de aprendizaje (en este caso, en los diferentes subsistemas de la EMS en México), ya que de ello dependió el grado de precisión de cada indicador, pues éstos, a su vez, también podían desagregarse en acciones más precisas.

El formato 2 muestra la relación entre las competencias disciplinares, los contenidos temáticos y los indicadores de aprendizaje para el ejemplo que se acaba de presentar.

Formato 2. Indicadores de aprendizaje en el contexto del currículo			
Perfil A. Se autodetermina y cuida de sí			
CG1. Se conoce y valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.			
CDLC1. Identifica, ordena e interpreta las ideas, datos y conceptos explícitos e implícitos en un texto, considerando el contexto en el que se generó y en el que se recibe.	6. Administra los recursos disponibles teniendo en cuenta las restricciones para el logro de sus metas.	CONTENIDO TEMÁTICO	INDICADORES DE APRENDIZAJE
Consolidación de los atributos	1	Planificación de un escrito	Organiza la redacción de un texto de acuerdo a su propósito comunicativo.
	1	Características de diversos tipos de texto.	Describe las partes que conforman las estructuras de los textos narrativos, expositivos, argumentativos y descriptivos.
	1		

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

A partir de esta información, el Comité Académico se concentró en la descripción de los indicadores de aprendizajes clave a evaluar en PLANEA EMS 2017, siguiendo la metodología anteriormente descrita. La importancia de seguir este proceso metodológico de manera cuidadosa se traduce en que se espera que los análisis estadísticos, tanto del piloteo del instrumento como de la aplicación definitiva, proporcionen puntajes que permitan verificar y validar no sólo la complejidad de cada reactivo y la confiabilidad de la prueba, sino también su pertinencia para la población objetivo de la evaluación.

Por esta misma razón, al término de esta tarea se procedió a la revisión de las descripciones de indicadores y se integró la tabla de indicadores de aprendizaje a evaluar, que fueron validadas por el Comité Académico en función de observar la suficiencia y exhaustividad de los indicadores de aprendizajes con respecto al perfil de egreso de la EMS y las competencias de Lenguaje y Comunicación y Matemáticas que los estudiantes deben desarrollar al término del nivel medio superior, descartando los que no se consideraron pertinentes. Todas estas acciones se dieron de manera colegiada y los acuerdos a los que se llegó fueron asentados por los coordinadores académicos de la prueba, mismos que resguardaron la documentación generada en esta fase del diseño de la prueba.

2.4. Tabla de contenidos e indicadores de aprendizaje

Para garantizar la confidencialidad, se presentan únicamente los contenidos adscritos a cada eje temático de cada disciplina, resguardando los indicadores de aprendizaje que constituyeron el instrumento de evaluación.

Tabla 3. Tabla de contenidos de Lenguaje y Comunicación PLANEA EMS 2107	
Eje temático	**Contenidos**
Manejo y construcción de la información	Modos discursivos
	Uso normativo de la lengua
	Estrategias de comprensión lectora
	Análisis de presupuestos ideológicos en textos
	Diversidad y selección de fuentes de información (como recurso)
	Análisis de valores éticos en las obras literarias y de arte
	Funciones de la lengua
	Proyecto de trabajo académico
	Análisis de las características de los trabajos académicos
	Análisis de medios y géneros informativos
	Resumen, paráfrasis, comentario, toma de notas
Texto argumentativo	Características del texto argumentativo
	Texto argumentativo escrito acerca de temas sobre diversidad
	Reseña crítica
	Debate, panel y disertación
	Estructura de textos argumentativos
	Ensayo
	Géneros periodísticos
	Operaciones textuales
	Análisis de textos literarios y argumentativos
Texto expositivo	Características del texto expositivo
	Propósito del texto argumentativo, narrativo, expositivo, descriptivo y dialógico
	Estructura de textos expositivos
	Análisis de textos expositivos cuya temática hable sobre situaciones sociales
	Estrategias de comprensión lectora de textos expositivos, artículos de divulgación científica y notas periodísticas
Texto literario	Texto narrativo en primera y tercera persona (biografía y diario)
	Análisis de obras recreativas
	Análisis de textos literarios
	Evaluación de textos literarios informativos y mensajes de medios de comunicación

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 4. Tabla de contenidos de Matemáticas PLANEA EMS 2107	
Eje temáticos	**Contenidos**
Sentido numérico y pensamiento algebraico	Ecuaciones
	Fracciones
	Jerarquía de operaciones
	Lenguaje algebraico
	Máximo común divisor
	Mínimo común múltiplo
	Números reales
	Polinomios
	Porcentajes
	Razones y proporciones
	Reducción de términos semejantes
	Sistemas de ecuaciones lineales
Cambios y relaciones	Distancia entre dos puntos
	Funciones
	Proporcionalidad

	Rectas y sus transformaciones
	Secciones cónicas
	Sucesiones
	Variación de parámetros
Forma, espacio y medida	Ángulos
	Semejanza
	Teorema de Pitágoras
Manejo de la información	Manejo de la información
	Medidas de dispersión
	Medidas de tendencia central
	Probabilidad

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 5. Indicadores de aprendizaje

Lenguaje y Comunicación	Matemáticas
130 indicadores de aprendizaje	102 indicadores de aprendizaje

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

2.5. Comité de Especificaciones

Si la Tabla de contenidos e indicadores de aprendizaje clave constituye el respaldo teórico y empírico del diseño de la prueba, las especificaciones de reactivos son la fase en la cual se operacionaliza el objeto de medida y se dictan las pautas y características para la construcción de los ítems (INEE, 2017). Esto a partir de la redacción de documentos que argumentan teórica y didácticamente los indicadores de aprendizaje. Por tanto, las especificaciones son el referente fundamental para la elaboración de los reactivos y son el vínculo entre éstos y el constructo, por lo que también sirven de criterio para determinar tanto la adecuada alineación de los ítems al marco conceptual establecido en el diseño de la prueba, como también para estandarizar los detalles técnicos para su elaboración.

Debido a su importante papel en la construcción de los reactivos que constituyen el instrumento de evaluación, el Comité de Especificaciones de PLANEA EMS 2017 debió estar conformado por profesionales expertos en las áreas disciplinares de Lenguaje y Comunicación y de Matemáticas en el nivel medio superior, especialistas en didáctica, formadores de docentes, expertos en el enfoque por competencias, docentes en ejercicio, investigadores o autores de libros. A partir de su experiencia, el Comité de Especificaciones fue capaz de argumentar los indicadores de aprendizaje de manera precisa y pudo visualizar el alcance de éstos, para determinar el formato de reactivo que pudiera representarlos a cabalidad.

2.5.1. Procedimiento metodológico para el desarrollo de las especificaciones

Para que el grupo colegiado encargado de la elaboración de las especificaciones pudiera realizar dicha tarea, fue necesario seguir un procedimiento metodológico dividido en tres fases de trabajo: capacitación, elaboración y revisión, con un Comité por cada área disciplinar. Los coordinadores académicos de la prueba se encargaron del diseño del Comité, de su capacitación, del asentamiento de los acuerdos y de la salvaguarda de los documentos generados, tanto digitales como impresos.
La primera fase de capacitación tuvo como objetivo adentrar al cuerpo colegiado en el marco general de PLANEA EMS 2017, a partir de:

1. La presentación de los objetivos de la prueba, 2. La exposición del marco de referencia para el desarrollo del instrumento, 3. La explicación de la delimitación conceptual del objeto de medida y 4. La presentación de las matrices de trabajo que justificaron los contenidos y los indicadores de aprendizaje clave, agrupados en la Tabla de competencias disciplinares básicas que consolidaron los atributos de las competencias genéricas, la Tabla de contenidos e indicadores de aprendizaje.

Todos los elementos anteriormente mencionados fungieron como insumos necesarios para que este Comité alineara sus referentes teóricos a las características y objetivos particulares de PLANEA EMS 2017.

Una vez en conocimiento de los lineamientos conceptuales y teóricos, se procedió a la explicación del procedimiento para la elaboración de especificaciones de reactivos, el cual se sustentó en los pasos establecidos por el INEE en sus *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* (2017), y en su *Manual técnico de especificaciones de reactivos* (2005). En esta segunda fase de la capacitación, los jefes de prueba y la subdirectora detallaron la estructura que establece el INEE para la construcción de las especificaciones de reactivos, la cual debe contener como elementos básicos:

> a) La definición operacional, de la cual puede identificarse el conocimiento o habilidad que se busca evaluar, el nivel de complejidad o proceso cognitivo, así como la situación contextual que le dé precisión; b) un ejemplo del reactivo o tarea evaluativa con la fundamentación de la respuesta óptima o correcta, así como de los distractores o respuestas parciales; y c) la bibliografía de apoyo (INEE, 2017: 19-20).

Una vez concluida la fase de capacitación, se procedió con la elaboración y los coordinadores académicos de la prueba presentaron ante el Comité el formato y la guía para la elaboración de las especificaciones, así como los ejemplos de redacción de especificaciones (ver Anexo III), materiales que reforzaron la comprensión de los elementos constituyentes de la estructura de las especificaciones. El objetivo fue que, en equipos de trabajo, los expertos participantes elaboraran una especificación a manera de ejercicio para ser analizada en plenaria.

A cada equipo se le presentó un esquema que contenía el campo disciplinar al que pertenecía la especificación, Lenguaje y Comunicación o Matemáticas, la competencia disciplinar y el atributo de la competencia genérica que ésta consolidaba, así como el contenido temático y su desglose analítico en un indicador de aprendizaje. Con estos estímulos, los especialistas argumentaron la teoría y didáctica del indicador de aprendizaje, cuidando su correspondencia con la Tabla de contenidos, que el reactivo muestra tuviera un nivel de dificultad adecuado con el grado escolar correspondiente, que los distractores fueran plausibles, que la redacción fuera clara, precisa y efectiva como herramienta para la construcción de ítems equivalentes, además de que estuviera libre de sesgo, y que se determinara la bibliografía de apoyo.

Cada equipo de trabajo tuvo la oportunidad de comentar, mediante plenaria, el trabajo de sus colegas y determinar si la especificación cumplía con la estructura básica solicitada por el INEE. Asimismo, cada grupo redactor modificó, de acuerdo con los comentarios, los elementos que ofrecían áreas de oportunidad, con la finalidad de afinar cada especificación para la última fase, la de revisión.

El procedimiento para la revisión de las especificaciones, cuyo protocolo se sustentó en el *Manual técnico de especificaciones de reactivos* (2005), del INEE, se llevó a cabo con base en el formato y la guía para la revisión de las especificaciones; además, se presentó un ejemplo de revisión, el cual fue analizado por el Comité de Especificaciones, quien, posteriormente a este proceso, estuvo en posibilidad de realizar un ejercicio de revisión de una especificación, a partir de los siguientes criterios:

1. IDENTIFICACIÓN DEL CONTENIDO A EVALUAR. Comprendió:

- El campo disciplinar al que pertenecía la especificación (Lenguaje y Comunicación o Matemáticas).
- La competencia disciplinar, la competencia genérica y el atributo, los cuales fueron determinados durante el Comité Académico de determinación de contenidos e indicadores de aprendizaje, en razón de que la competencia disciplinar contribuía a la consolidación del atributo de la competencia genérica y se relacionaban con los planes y programas de los subsistemas de la EMS.
- El contenido para la elaboración del reactivo (el tema general del contenido a evaluar).
- El indicador de aprendizaje.

2. DESCRIPCIÓN GENERAL DEL CONTENIDO A EVALUAR. Incluyó:

- La delimitación del contenido que se debía cubrir, a partir de una elaboración teórica que debía seguir el método deductivo, con la finalidad de argumentar teóricamente el aspecto más relevante del aprendizaje a evaluar, definido como el indicador.
- La interpretación del sentido del contenido, la cual detallaba lo que se pretendía que el estudiante pudiera realizar a partir del proceso de enseñanza aprendizaje que recibió en su educación formal, a partir de las siguientes preguntas: ¿Qué habilidades necesita el alumno para lograr el aprendizaje del contenido a evaluar? ¿Qué dificultades implica el aprendizaje o la evaluación de este contenido? ¿Qué dificultades manifiestan los estudiantes en este contenido? ¿Cuáles son los errores más comunes en el estudio de este contenido?
- La importancia del contenido en función de la consolidación del atributo y la competencia disciplinar, es decir, se debía argumentar qué tan sustancial era dicho contenido para la consolidación de los atributos descritos en el apartado "Identificación del contenido a evaluar", a partir de la siguiente pregunta orientadora: "¿Cómo apoya este contenido en la consolidación del atributo *X* de la competencia genérica *Y*, tomando en consideración la competencia disciplinar *Z* en la que se ubica?
- La función del contenido en el contexto de la disciplina, es decir, se debía evidenciar la importancia del contenido para generar nuevos aprendizajes, contextualizándolo en el área disciplinar que lo englobaba (Lenguaje y Comunicación o Matemáticas), a partir de dos preguntas clave: ¿Qué importancia tiene este contenido para el desarrollo de habilidades o competencias del campo disciplinar? ¿Cómo se vincula el aprendizaje a evaluar con otros aprendizajes o contenidos?

3. NIVEL DE DOMINIO. Integró:

- La descripción del nivel de dominio esperado por parte del estudiante, que integró las temáticas necesarias e indispensables que el estudiante debía aplicar en la resolución del reactivo propuesto y justificó la demanda cognitiva que se le solicitó.

- Los conocimientos y habilidades que el estudiante requirió para contestar correctamente el reactivo, las cuales se precisaron a través de un listado descriptivo de éstos.
- Las actividades cognoscitivas involucradas en la respuesta, que son los procesos intelectuales que se ponen en marcha cuando el estudiante responde al reactivo. En este sentido, dichas actividades cognitivas estuvieron enmarcadas por la *Taxonomía de Bloom*, la cual surgió, precisamente, de la necesidad de generar una estructura que explicitara los niveles de abstracción en la elaboración de tareas evaluativas.

4. PLANTILLA. Circunscribió las características básicas y generales para la elaboración de los ítems equivalentes, a partir de un esquema que detalló:

- La estructura de la base del reactivo (las características que debía contener el enunciado, instrucción, pregunta o consigna y, si fuera el caso, los elementos de apoyo).
- La estructura y descripción de la opción correcta, así como la estructura y descripción de las opciones consideradas incorrectas.
- Las fuentes de consulta utilizadas para la elaboración de dicha plantilla (listadas en formato APA).

5. REACTIVO MUESTRA. Reactivo-tipo en el que se operacionalizó la estructura, las características detalladas en el punto anterior y se desglosó:

- La estructura de la base del reactivo.
- La respuesta correcta y los distractores, cuidando que no hubiera más de una respuesta correcta o respuestas parcialmente correctas.
- Las fuentes consultadas (listadas en formato APA).

6. FUENTES DE CONSULTA. Incluyó la bibliografía, hemerografía o la ciberografía en la que se sustentó tanto la investigación teórica como la didáctica, enlistada a partir de las normas APA.

Pero además de estos elementos técnicos y de contenido, también fue importante la revisión de los criterios relacionados con el sesgo y la diversidad cultural, pues el equipo de Felipe Martínez Rizo ya había señalado la importancia de este aspecto en su estudio de 2015, al detallar que algunos problemas "de naturaleza lingüística (estructura sintáctica, uso de términos infrecuentes o inexistentes en el español de México, fraseo poco claro y con información imprecisa)" (Martínez, 2015: 13), debían tener una revisión detallada debido al contexto multicultural que representa nuestro país y bajo la premisa de no vulnerar la equidad de la evaluación al poner en desventaja a una población escolar con respecto a otra. Las consecuencias de que las evaluaciones a gran escala presenten sesgo cultural decantan en conclusiones y toma de decisiones en política educativa poco eficientes en el contexto educativo mexicano.

En contraste con el funcionamiento diferencial de los ítems de carácter psicométrico, el sesgo cultural es un elemento que funge como barrera para que los estudiantes demuestren habilidad en el constructo que se desea medir (Gipps *et al.*, 2010), por ello se relaciona directamente con la dimensión ética de justicia y equidad, ya que se produce a partir de condiciones culturales, socioeconómicas, lingüísticas, étnicas y de género. De acuerdo con la literatura especializada, el sesgo se refiere:

> a la injusticia derivada de uno o varios ítems del test al comparar distintos grupos, que se produce como consecuencia de la existencia de alguna característica del ítem o del contexto de aplicación del test que es irrelevante para el atributo medido por el ítem (Gómez-Benito *et al.*, 2010: 76).

Por ello, PLANEA EMS 2017 cuidó desde el diseño de la prueba este aspecto (al incluir a docentes de bachillerato comunitario y bachillerato intercultural). En la etapa de elaboración de especificaciones también se atendió este rasgo procurando, en primer lugar, contar con representación de los diferentes subsistemas de la EMS a partir de la invitación extendida a los profesores que formaban parte de éstos. En segunda instancia, se definieron algunos criterios ligados con el sesgo cultural, étnico, lingüístico y de género que los constructores de las especificaciones debieron respetar e integrar a su labor, mismos que posteriormente fueron afinados en la fase de validación de los ítems:

- De naturaleza lingüística: procurar la construcción de estructuras sintácticas claras, facilitando un orden lógico; evitar el uso de términos poco frecuentes o inexistentes en el español mexicano, así como el uso de regionalismos que pudieran modificar su semántica en las distintas entidades federativas, grupos sociales o étnicos; evitar el uso de información imprecisa en las formulaciones de las especificaciones y apegarse, ante todo, al español estándar mexicano.
- De registro: procurar alinearse a un registro lo más objetivo posible.
- De contexto: evitar el uso de situaciones que resultasen familiares para determinados grupos poblacionales en detrimento de otros, o bien, eludir escenarios que no fueran referenciales de la realidad de los estudiantes de la EMS.
- De contenido: identificar y suprimir prejuicios de todo tipo (raciales, ideológicos, de género, religiosos, de preferencia sexual, económicos), así como presupuestos ideológicos que pudieran ser ofensivos o denigrantes para ciertos grupos poblacionales.

Con estos criterios, se redujeron las posibles fuentes de varianza irrelevante al constructo y garantizar la accesibilidad del instrumento al total de la población sujeto de la evaluación, accesibilidad que: "debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir" (AERA-NCME-APA, 2014: 49). Con estos criterios, aunados a los de carácter técnico, la Dirección de Evaluaciones Nacionales de Resultados Educativos del INEE procuró respetar el principio de la evaluación con equidad, a partir del diseño de sus evaluaciones.

2.6. Ejemplos de especificaciones

Para efectos de esta descripción, se presentan los documentos de las especificaciones comprimidas y, posteriormente, se muestran ejemplos de los reactivos derivados de estas especificaciones.

Formato 3. Especificación de Lenguaje y Comunicación

DIRECCIÓN GENERAL DE EVALUACIONES DE RESULTADOS EDUCATIVOS
DIRECCIÓN DE EVALUACIONES NACIONALES DE RESULTADOS EDUCATIVOS
EVALUACIÓN DE LA EDUCACIÓN MEDIA SUPERIOR
FORMATO PARA LA ELABORACIÓN DE LAS ESPECIFICACIONES DE REACTIVOS

1. IDENTIFICACIÓN DEL CONTENIDO A EVALUAR

a. Campo disciplinar: Lenguaje y Comunicación.
b. Competencia disciplinar: 7. Valora y describe el papel del arte, la literatura y los medios de comunicación en la recreación o la transformación de una cultura, teniendo en cuenta los propósitos comunicativos de distintos géneros.
c. Competencia genérica: 2. Es sensible al arte y participa en la apreciación e interpretación de sus expresiones en distintos géneros.
d. Atributo: 2. Experimenta el arte como un hecho histórico compartido que permite la comunicación entre individuos y culturas en el tiempo y el espacio, a la vez que desarrolla un sentido de identidad.
e. Contenido para la elaboración del reactivo (tema general del contenido a evaluar): Lectura de textos literarios.
f. Indicador de aprendizaje: Identificar la intención comunicativa del texto literario.

2. ¿QUÉ FUNCIÓN TIENE EL CONTENIDO EN LA ENSEÑANZA DEL LENGUAJE Y LA COMUNICACIÓN?

La lectura de textos literarios contribuye al desarrollo de habilidades lingüísticas como escuchar, hablar y escribir (Serratos, s/f) porque leer textos de diferentes géneros permite, por una parte, conocer distintos temas que son de interés general para el ser humano y, a su vez, las formas en que son presentados; por otra parte, también permite identificar las maneras en que se puede expresar una opinión y los comentarios sobre el contenido; desarrollar la imaginación, o bien, producir textos complejos que integren al unísono procesos como el análisis y la crítica.

Todo mensaje tiene un propósito, es decir, se expresa un mensaje para lograr una finalidad. Comunicar significa intercambiar mensajes. Éstos están conformados por diversos tipos de información como conceptos, sentimientos u opiniones. Los textos literarios guardan dentro de sí un mensaje que el autor busca transmitir.

Los seres humanos expresan mensajes orales y escritos y la intención de quien los expresa representa un rasgo del carácter racional de la conciencia, es decir, el enunciador del mensaje es consciente de algo y busca ir más allá del acto comunicativo en sí mismo para hacer trascender (Serratos, s/f). Por lo anteriormente expuesto, es importante que los estudiantes aprendan a *identificar la intención comunicativa del texto literario,* pues éste en sí aborda infinidad de temas que se ponen al alcance del lector, no sólo para disfrutarlos, sino también para dotarlos de cultura o para que desarrollen otras capacidades cognitivas, emocionales y sociales. La actividad de identificar la intención comunicativa requiere de diversas estrategias de lectura y de la experiencia del lector, también es una acción básica en la que se integran otras operaciones mentales.

El pensamiento se construye con palabras y el manejo coherente de éstas se aprende leyendo. Saber pensar implica saber organizar los pensamientos. Éste último saber es el punto de partida para generar conocimiento. Así pues, para aprender es importante saber leer. Por ejemplo, quien domina la lengua oral y escrita tendrá altas posibilidades de comprender una gran cantidad de conocimientos.

De acuerdo con lo anterior, Rafael Ruiloba (2013) considera que:

> por medio de la literatura los jóvenes deben aprender a pensar, por tanto, deben aprender a escribir ensayos y, por tanto, deben leer ensayos cortos y saber argumentar. Así se relaciona la lectura literaria de ensayos con la escritura. El ensayo es el género literario que relaciona el lenguaje con la realidad, porque es un género no ficticio.

De este modo, identificar la intención comunicativa del texto literario es fundamental para que el alumno pueda tener un amplio criterio, ser juicioso y reflexivo ante la diversidad de textos que se le presenten en el contexto escolar, además, se fomenta la comprensión de la lectura la cual es una de las habilidades más importantes para la interpretación de un texto sin importar su extensión. Comprender lo que se lee es una habilidad presente no sólo en la lectura recreativa sino también en la lectura para estudiar e incorporar nuevos conocimientos.

Finalmente, la lectura de textos literarios también apoya la competencia de saber autodeterminarse y cuidar de sí mismo. Durante la lectura de obras literarias, los jóvenes en la educación media superior pueden identificarse con historias que son afines a su comportamiento o forma de pensar, pero también los acerca a dimensiones de su identidad que posiblemente son desconocidas. La lectura es una posibilidad para el estudiante de autoconocerse por medio de historias ficticias, contribuye a la constante búsqueda de su ser, de conocer y reconocerse en el otro (empatía por ciertos personajes dentro de una historia), para vincularse de manera social, histórica y cultural.

3. ¿QUÉ CONOCIMIENTOS Y HABILIDADES SE REQUIEREN PARA CONTESTAR EL REACTIVO CORRECTAMENTE?

Las habilidades que el alumno requiere para identificar la intención comunicativa están relacionadas con la comprensión lectora:

- Conocer el proceso comunicativo en los textos literarios.
- Conocer la función de la lengua que prevalece en los textos literarios.
- Conocer las intenciones comunicativas de los textos literarios.
- Conocer los diferentes géneros literarios, su estructura y elementos.
- Conocer estrategias y tipos de lectura para abordar los textos literarios.
- Aplicar diferentes tipos de lectura en el texto literario.
- Aplicar estrategias de lectura para comprender e interpretar el texto literario.
- Reconocer el emisor, receptor, contexto y referente en el texto literario.
- Ubicar la función de la lengua que prevalece en el texto literario.
- Identificar el género, su estructura y elementos en el texto literario.
- Inferir la intención comunicativa en el texto literario.

4. FUENTES DE CONSULTA

Araújo, N. y Delgado, T. (2003). *Textos de teorías y críticas literarias*. (Del formalismo a los estudios postcoloniales). México: Universidad Autónoma Metropolitana / Universidad de la Habana.
Camarero, M. (s/f). *Introducción al comentario de textos*. Madrid: Castalia.
COLBACH. Colegio de Bachilleres. (2012). *Intención comunicativa de los textos*. México: Autor.
Garrido, M. (2004). *Nueva introducción a la teoría de la literatura*. Madrid: Síntesis.
Pozuelo, J. (1988). *La teoría del lenguaje literario.* Madrid: Cátedra.
Searle, J. (1980). *Actos de habla*. Madrid: Cátedra.
Serratos, M. [Ponencia]: *Importancia de la lectura en la adquisición de conocimientos*. Recuperado el 26 de abril de 2019 de: http://www.cch-oriente.unam.mx/areas/talleres/Ponencias%20XIII%20Encuentro%20Profs%20PDF/Ma.%20Refugio%20Serratos%20Glez.pdf
Solé, I. (1992). *Estrategias de lectura*. Barcelona: Graó.

Formato 4. Especificación de Matemáticas

DIRECCIÓN GENERAL DE EVALUACIONES DE RESULTADOS EDUCATIVOS
DIRECCIÓN DE EVALUACIONES NACIONALES DE RESULTADOS EDUCATIVOS
EVALUACIÓN DE LA EDUCACIÓN MEDIA SUPERIOR
FORMATO PARA LA ELABORACIÓN DE LAS ESPECIFICACIONES DE REACTIVOS

1. IDENTIFICACIÓN DEL CONTENIDO A EVALUAR

a. Campo disciplinar: Matemáticas.
b. Competencia disciplinar: 1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
c. Competencia genérica: 5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.
d. Atributo: 1. Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
e. Contenido para la elaboración del reactivo (tema general del contenido a evaluar): Funciones.
f. Indicador de aprendizaje: Realizar operaciones con funciones.

2. ¿QUÉ FUNCIÓN TIENE EL CONTENIDO EN LA ENSEÑANZA DE LAS MATEMÁTICAS?

Aprender a operar con funciones permite movilizar y reforzar los conocimientos previos, por ejemplo, las operaciones algebraicas, para conectarlos con los nuevos, o comprender las funciones transcendentales, como el logaritmo o la exponencial, en su forma básica y, a partir de ellas, conocer cómo es posible construir nuevas funciones del mismo tipo.

En general, operar con las funciones exige que los estudiantes dejen de verlas como un proceso o procedimiento a seguir, para mirarlas como factores sobre los cuales se puede operar. Este cambio en su concepción sobre la función resulta de vital importancia para comprender conceptos posteriores como derivadas e integrales. Además, permite comprender diversos conceptos de precálculo y cálculo, así como de matemáticas financieras y estadística.

El nuevo conocimiento viene a fortalecer el proceso de razonamiento siempre que esté en constante uso, por ello la importancia de resolver operaciones de forma generalizada, pero con sentido de utilidad en los problemas de su entorno o con la finalidad de hacerlo más significativo.

3. CONOCIMIENTOS Y HABILIDADES REQUERIDOS PARA CONTESTAR CORRECTAMENTE EL REACTIVO:

- Dominar las operaciones aritméticas y algebraicas.
- Usar la notación de funcional.

4. FUENTES DE CONSULTA

Cantoral, R. y Montiel, G. (2001). *Funciones. Visualización y pensamiento matemático*. México: Pearson Educación.

Cuesta, V., Soto , E., y Lezama, M. (2008). *Matematicas IV.* Mexico: Book Mart.

Soto, J., Álvarez, A., Otero, G., Lampón, J. y Grinda, J. (2006). *Bachillerato Intercultural. Matemáticas Aplicadas 1. Guía del docente*. México: Secretaría de Educación Pública.

Ortiz, F., Ortiz, F., y Ortiz, F. (2015). *Matemáticas.* Mexico: Patria

3. Desarrollo del instrumento de evaluación

En la segunda fase de la construcción de un instrumento de evaluación se concretan los referentes teóricos y conceptuales referidos en la etapa del diseño. Esta es una fase primordial que inicia con la construcción de los reactivos, la revisión técnica y de sesgo cultural y lingüístico de los mismos; engloba las tareas de aplicación de la versión piloto: su integración, revisión y reproducción impresa; la aplicación; el análisis de resultados y el estadístico de los ítems; y, finalmente, la integración de la prueba definitiva con la selección de reactivos a partir de criterios estadísticos, la distribución de los bloques, la edición, el armado y la impresión de la prueba final.

Todos estos pasos tienen procedimientos metodológicos específicos y tanto la construcción de reactivos como su revisión técnica y de sesgo cuentan con la participación de cuerpos colegiados de expertos de la educación media superior (EMS) desde la docencia y la investigación. Estos profesionales garantizan el rigor técnico y la calidad del contenido de los reactivos de la prueba PLANEA EMS 2017, y contribuyen a reducir al mínimo el sesgo cultural. A continuación se describe el procedimiento que estos equipos de trabajo siguieron para la elaboración de los productos encomendados a cada uno de los Comités.

3.1. Comité de Elaboración de Reactivos

Debido a que los ítems representan de manera directa el universo de medida de la prueba y son el elemento al que se enfrenta el estudiante, para PLANEA EMS 2017 fue importante que los integrantes del Comité fueran, además de profesionales expertos en las áreas de Lenguaje y Comunicación y Matemáticas en el nivel medio superior, preferentemente docentes frente a grupo, ya que su conocimiento de los errores más comunes de los estudiantes fue determinante para la elaboración de distractores plausibles y, en general, una herramienta para cuidar que el nivel de dificultad de los reactivos correspondiera con el del nivel educativo a evaluar. Del mismo modo, también se procuró que este Comité tuviera una amplia y heterogénea representatividad de los diferentes subsistemas de la EMS, con la finalidad de prevenir el sesgo y procurar la equidad en la evaluación.

El Comité de Elaboración de Reactivos fue el tercer cuerpo colegiado que participó en la construcción de PLANEA EMS 2017 y tuvo a su cargo efectuar las operaciones indicadas en las especificaciones, para generar ítems equivalentes para cada indicador de aprendizaje de la Tabla de contenidos. Para concretar de manera satisfactoria esta tarea, el Comité se apegó a la estructura y a los criterios establecidos en cada especificación y, con ello, garantizó que el instrumento de evaluación efectivamente midiera el constructo propuesto y argumentado desde la fase de diseño. Se debe tener en cuenta que PLANEA EMS 2017 se valió sólo de ítems de opción múltiple, en el entendido de que: "Por distintas razones, especialmente por su versatilidad y características psicométricas [...] pueden ser utilizados en distintos contextos y con diferentes propósitos; sirven para evaluar conocimiento factual (puramente memorístico), habilidades intelectuales de alto orden, o disposiciones actitudinales y valorativas" (INEE, 2005b: 13). Por esta razón, este cuerpo colegiado se guío con los lineamientos técnicos propuestos por el Instituto Nacional para la Evaluación de la Educación (INEE) en el *Manual técnico. Construcción de reactivos* (2005), el cual detalla las normas técnicas a las que se deben ceñir los ítems de opción múltiple, para que tengan congruencia con las especificaciones y, por ende, con el universo de medida que éstas representan.

3.1.1. Procedimiento metodológico para la elaboración de reactivos

Para desarrollar las habilidades que dotaron al Comité de Elaboración de Reactivos de las bases técnicas y estilísticas para la ejecución de su tarea, los jefes de prueba y la subdirectora del área diseñaron sesiones teórico-prácticas tanto de trabajo individual como colegiado. En estas sesiones se capacitó a los integrantes del Comité en la elaboración de reactivos de opción múltiple, con la presentación de ejemplos y la elaboración de ejercicios basados en los criterios técnicos del INEE, para, de esta manera, acordar las características que debían tener tanto la base del reactivo como las opciones de respuesta correcta y los distractores, además de evitar en lo posible el sesgo cultural.

El primer paso en la metodología del Comité fue la introducción al PLANEA EMS 2017, para que los integrantes del cuerpo colegiado conocieran los elementos que sustentaron la prueba: el marco de referencia, los propósitos y la población objetivo; además de los referentes utilizados para determinar los contenidos e indicadores de aprendizaje clave asentados en la Tabla de contenidos. El objetivo fue que los expertos comprendieran el propósito de la prueba, sus alcances y limitaciones, así como la relevancia de los contenidos en la consecución de dichos objetivos.

Un punto clave en el desarrollo del Comité fue la explicación detallada de las características de las especificaciones, ya que éstas representaban el sustento teórico, conceptual y técnico de los ítems. El *Manual técnico. Construcción de reactivos* señala la importancia de realizar este proceso, ya que: "La idea es conocer con mucha precisión la interpretación de los contenidos a evaluar, su importancia y función curricular, así como la actividad cognoscitiva involucrada en el reactivo a contestar" (INEE, 2005b: 18). De lo anterior se desprende que con la comprensión cabal de estos elementos, los reactivos elaborados serán congruentes con su especificación y representarán adecuadamente el indicador de aprendizaje. Por lo tanto, en el análisis de las especificaciones, los encargados de la construcción de reactivos respondieron las siguientes preguntas orientadoras para comprender el sentido de cada especificación y sus requerimientos:

- ¿Qué es lo que se pretende evaluar? Para entender el sentido del contenido a evaluar.
- ¿Qué actividad intelectual se está evaluando con este reactivo? Para determinar la habilidad cognoscitiva que se activa al momento de responder el ítem.
- ¿En qué medida la plantilla corresponde a la esencia del contenido y a la habilidad intelectual que se desea evaluar? Para observar la congruencia entre la plantilla y los demás elementos de la especificación.
- ¿El reactivo corresponde a las características y plantilla descritas en la especificación (en cuanto a contenido y habilidades requeridas)? Para analizó la congruencia, esta vez entre el reactivo muestra y su plantilla (INEE, 2005b: 18).

Además del análisis formal de las especificaciones, se presentaron el formato de elaboración de reactivos y los criterios para su elaboración señalados en el *Manual técnico. Construcción de reactivos* (INEE, 2005b), insumos necesarios para concretar estos aspectos teóricos en ejercicios prácticos.

- **Criterios técnicos y de estilo para la elaboración de reactivos**

1. De la congruencia con su especificación
    a. Tanto la base del reactivo como los distractores deben redactarse de conformidad estricta con su correspondiente especificación.

b. Ningún ítem debe ser más o menos difícil de lo que señale su especificación.
c. Se podrán utilizar los libros de texto, guías y otros materiales didácticos oficiales como insumo para la construcción de ítems. Sin embargo, debe evitarse la creación de reactivos idénticos a los que aparecen en dichos recursos.
d. En los casos en que resulte necesario cambiar la dificultad del reactivo, porque así lo indique el análisis de la prueba empírica, la modificación se efectuará incrementando la dificultad de las opciones de respuesta y nunca formulando la base de la pregunta de manera más ambigua o mediante el uso de vocabulario más sofisticado o de situaciones más complejas.

2. Del contenido

El contenido de los reactivos deberá:

a. Reflejar el currículo nacional [en el caso de PLANEA EMS 2017, reflejar el cruce entre el MCC y los planes y programas de estudio].
b. Centrarse en los contenidos curriculares que, por su importancia y frecuente implementación dentro del aula, sean más significativos en el nivel escolar al que se dirige la prueba.
c. Evaluar lo que es esencial y consecuente con un determinado aprendizaje, evitando tener que recurrir a conocimientos adquiridos fuera del salón de clases.
d. Considerar los contenidos, el vocabulario y los procesos de aprendizaje previos, pero sólo para lograr el andamiaje preciso con el contenido a evaluar y partir de lo que el alumno ya conoce.
e. Evitar requerir del estudiante el uso de una estrategia en particular para resolver los problemas que se le planteen.
f. Evitar entrar en detalles del conocimiento que resulten demasiado específicos, así como emplear información basada en opiniones.
g. Procurar que cada reactivo constituya un estímulo y un reto intelectual [para los estudiantes].

3. Del uso del lenguaje

El lenguaje de los reactivos deberá:

a. Ser sencillo y acorde al nivel escolar al que se dirige la prueba.
b. Tener calidad ortográfica y de redacción.
c. Estar libre de expresiones idiomáticas locales que dificulten su comprensión.
d. Evitar todo tipo de sesgo (cultural, social, de género, etcétera); es decir, vocabulario o cualquier tipo de representación que ofenda a un grupo de estudiantes en particular o facilite la identificación de la respuesta correcta o dificulte contestar correctamente.
e. Contemplar el uso de nombres propios representativos de la diversidad cultural de México. Se preferirán nombres cortos que sean fáciles de leer.

4. De la redacción

La redacción de un reactivo deberá:

a. Tener oraciones breves y sintaxis sencilla.
b. Especificar claramente lo que se espera como respuesta y ayudar al estudiante a enfocar su atención en la tarea a realizar.
c. Considerar el uso de la organización en viñetas para el caso específico de instrucciones complejas.
d. Contemplar un tiempo mínimo de lectura, lo cual exige, por lo tanto, precisión, sencillez y claridad de ideas.

e. Utilizar una combinación semántica y sintáctica simple.
f. Ser siempre afirmativa.
g. Usar verbos en imperativo al inicio de la instrucción o instrucciones.
h. Establecer consignas directas y concisas.
i. Construir la base de cada ítem en forma de pregunta o de enunciado incompleto.
j. Redactar cada ítem de manera independiente de los demás reactivos.
k. Formular el reactivo en forma positiva evitando la palabra "NO" en la base de la pregunta.
l. Evitar mezclar el modo condicional y el modo interrogativo en una misma oración. Por ejemplo: "¿Cuál sería el resultado de sumar 25 más 32, si 32 es un número negativo?".
m. Incluir todos los elementos de información necesarios para responder cada reactivo.

5. De los materiales de apoyo (ilustraciones, gráficos, tablas, etcétera)
Los materiales de apoyo deberán:

a. Ser concisos.
b. Ser conceptualmente correctos.
c. Ser claros y simples, con el mínimo de distractores o información irrelevante.
d. Cumplir su función de apoyo y nunca "engañar" a los estudiantes para que se confundan o seleccionen una respuesta incorrecta.
e. Ser realistas y cercanos a la cotidianidad de los niños o adolescentes, particularmente en el caso de representaciones gráficas como dibujos o fotografías.
f. Ser atractivos y acordes al nivel escolar a evaluar.

6. De formato

a. Se evitará usar margen justificado.
b. La fuente tipográfica a utilizar será Arial, 12 puntos.
c. Las imágenes deberán tener formato JPG.

7. De las preguntas de opción múltiple

a. En promedio, el reactivo debe requerirle al estudiante un minuto para su respuesta. Ocasionalmente podrá utilizar dos o tres minutos.
b. Cada reactivo deberá contener una base del reactivo (en formato de pregunta, enunciado incompleto o afirmación) y cuatro posibles respuestas (tres plausibles, pero sólo una considerada correcta).
c. La base del reactivo deberá presentar información suficientemente clara, a fin de que el estudiante sepa qué hacer antes de leer las opciones de respuesta.
d. En la base del reactivo se evitará incluir información irrelevante.
e. Cuando el reactivo lo requiera (por ejemplo, cuando incluya textos o figuras complementarias), podrá acompañarse de instrucciones adicionales a la base de la pregunta.
f. Los distractores deberán reflejar errores comunes o malos entendidos, concepciones simples o ingenuas, u otro tipo de error, de tal forma que las respuestas correctas demuestren lo que los alumnos realmente saben o pueden hacer.
g. Los distractores no tendrán la intención de confundir al estudiante, por lo que no deberán ser variantes cercanas a la respuesta correcta.
h. Todos los distractores deberán tener congruencia gramatical con la pregunta.
i. No se deberán utilizar alternativas como: "todas las anteriores", "ninguna de las anteriores", "las dos de arriba".

j. El diseño de las alternativas no deberá inducir a que los estudiantes descarten o eliminen alguna opción de respuesta, o a que identifiquen la respuesta correcta debido a características superficiales y triviales de las otras alternativas, o debido a la complejidad sintáctica y conceptual de las mismas.
k. Los reactivos no deberán permitir que las respuestas correctas se obtengan mediante procedimientos incorrectos.
l. Preferentemente, las alternativas de respuesta deberán tener la misma longitud y formato, así como ser sintáctica y semánticamente paralelas. En caso de que esto no sea posible, podrá haber dos pares de alternativas con las mismas características (ejemplo: dos respuestas cortas y dos largas).
m. Las letras A, B, C y D deberán ser utilizadas para las alternativas de respuesta y no como encabezados o viñetas en el reactivo.
n. Las alternativas de respuesta deberán ordenarse considerando su longitud, cantidad o cronología.
o. En las especificaciones de reactivos la primera opción deberá ser siempre la respuesta correcta. Deberá, además, estar indicada con un asterisco.
p. Ningún ítem debe proporcionar determinantes específicos; es decir, pistas que indiquen la respuesta correcta o que invaliden ciertas opciones incorrectas que se ofrecen a los estudiantes.

A continuación se presentan los determinantes específicos más comunes [situaciones a evitar]:

- Se proporcionan claves contextuales, las cuales permiten identificar que una opción no puede ser correcta o, en su caso, que se trata de la respuesta correcta. Por ejemplo, la base del reactivo se refiere a un tema determinado; tres alternativas hacen referencia al mismo tema, y la restante, que es incorrecta, se refiere a un tema distinto.
- Se proporcionan claves gramaticales, las cuales indican que una opción es la correcta o que no lo es. Por ejemplo, la base informativa de un ítem demanda que la respuesta correcta sea un sustantivo con número singular; sin embargo, salvo la opción correcta, las demás opciones presentan sustantivos cuyo número es plural.
- La longitud de una opción es considerablemente más larga o más corta que las demás.
- Se emplean determinantes como “nunca”, “siempre” o “algunas veces”.
- Se presentan asociaciones evidentes, opciones absurdas o indicadores semánticos muy generales o específicos.

8. De las pruebas de Matemáticas
a. Con base en la pregunta, y para el caso particular del uso de números de más de tres dígitos enteros (que no refieran años calendario), se deberán utilizar espacios para agrupar órdenes y clases numéricas. Por ejemplo: 26 000.
b. Se podrán utilizar abreviaturas de medidas estándar; sin embargo, cuando pueda haber una posible confusión, será necesario evitarlas. Por ejemplo: pulgada (para sustituir la abreviatura “pul”).
c. Las variables deberán usar tipografía itálica y mayúscula.
d. La variable X no deberá usarse en los reactivos donde el símbolo “x” se utilice para indicar una multiplicación.
e. Toda fracción quedará representada mediante una línea horizontal para separar al numerador del denominador. Por ejemplo: $\frac{1}{2}$

f. Los reactivos no deberán evaluar definiciones de vocabulario.
g. Las características de las cifras y símbolos dependerán del eje y grado escolar, tomándose las correspondientes de los libros de texto.
h. En la solución de problemas escritos o en contexto, las medidas utilizadas deberán ser cercanas a la realidad. Nunca se deberán utilizar situaciones absurdas.

9. De las pruebas de Lenguaje y Comunicación
   a. Se deberán usar instrucciones adicionales a la base del reactivo: una para la acción de lectura y otra para la actividad a desarrollar a partir del texto.

Los textos de lectura tendrán las siguientes características:
   b. Deberán ser tomados de libros, revistas o periódicos de reconocido prestigio. Siempre será necesario referir la fuente de consulta.
   c. El uso de fragmentos o adaptaciones estará sujeto a las necesidades que se deriven de la prueba en general, o del reactivo en particular, y deberá ser autorizado por el Coordinador [de la prueba].
   d. Los diversos textos que se seleccionen deberán contener información fidedigna y de interés para los estudiantes del nivel educativo a evaluar (INEE, 2005b: 27-32).

- **Estructura del reactivo**

Además de los criterios anteriores, otro elemento de importancia fundamental es la definición y seguimiento de una estructura formal para los reactivos, ya que esta tarea permite la construcción estandarizada del banco de ítems para los instrumentos de evaluación, al mismo tiempo que apoya el ordenamiento de los componentes básicos que los reactivos deben contener y dota de claridad al proceso. En el caso de PLANEA EMS 2017, el Comité de Elaboración de Reactivos ciñó la construcción de éstos a la estructura formal que el INEE ha establecido en sus documentos normativos, que se divide en los siguientes apartados:

1. DATOS GENERALES
   - Nombre del redactor del reactivo.
   - Campo disciplinar al que pertenecía la prueba.
   - Número de entrega del reactivo.
   - Nombre del coordinador académico.
   - Identificador único del reactivo.

2. IDENTIFICACIÓN DEL CONTENIDO A EVALUAR
   - Competencia disciplinar.
   - Competencia genérica.
   - Atributo.
   - Contenido temático.
   - Indicador de aprendizaje como aparecía en la especificación.
   - Identificador único de la especificación.

3. REACTIVO COMPLETO
   - Estructura de la base del reactivo: estímulo en forma de pregunta.
   - Respuesta correcta: opción que respondía correctamente a la base de la pregunta.

- Distractores: opciones incorrectas pero plausibles para los estudiantes con niveles de logro insuficientes.
- Fuentes de consulta: referencias que fungieron como apoyo teórico o didáctico en la construcción del reactivo, o bien, fuentes de donde se extrajeron las figuras y textos adicionales.

Este apartado también pudo contar con componentes adicionales, los cuales dependían de la naturaleza del reactivo. Estos elementos fueron:

- Instrucciones: dirigidas a los textos y a las figuras adicionales.
- Figuras adicionales: elementos gráficos (dibujos, diagramas, gráficas, fórmulas, etcétera) que el estudiante debía analizar para responder la pregunta.
- Textos adicionales: elementos que acompañaban a las figuras adicionales o escritos completos que el estudiante debía leer para responder el reactivo (INEE, 2005b: 13).

4. JUSTIFICACIÓN DE LAS OPCIONES DE RESPUESTA. Este apartado comprendió:

- Justificación de la respuesta correcta: argumentación del porqué la opción denominada como "correcta" respondía de manera adecuada a la base del reactivo.
- Justificación de los distractores: argumentación del porqué las demás opciones se debían tomar como elementos incorrectos para responder la base del reactivo. Estos distractores podían estar sustentados con base en las experiencias docentes acerca de los errores más comunes de los estudiantes en cada contenido curricular.

Una vez puntualizados los criterios técnicos, de estilo y la estructura del reactivo, el Comité de Elaboración de Reactivos desarrolló tres reactivos para cada especificación. En un segundo momento, dichos reactivos se asignaron a otro grupo de trabajo para su revisión, a fin de verificar cuáles elementos no eran congruentes con los lineamientos técnicos, de contenido o de sesgo cultural. Posteriormente, los elaboradores realizaron los ajustes a los reactivos atendiendo a las observaciones recibidas.

Posteriormente, con la Lista de cotejo para la revisión de reactivos, documento que sintetiza las características técnicas y de estilo a las que cada reactivo debe responder se revisaron los reactivos. Para trabajar con este documento, los jefes de prueba y la subdirectora detallaron el procedimiento para la revisión mediante la Lista de cotejo, el cual constó de dos criterios específicos: 1) Cada uno de los elementos del reactivo debió apegarse a la especificación correspondiente. 2) Las características del reactivo debieron apegarse a las *Normas para la construcción de reactivos* (INEE, 2005b: 21). Para apoyar este proceso, el cuerpo colegiado respondió a las siguientes interrogantes:

- Revisión de la congruencia entre la especificación y el reactivo: ¿El ítem es congruente con la especificación? ¿El ítem es congruente con el contenido? ¿El ítem es relevante?
- Revisión de la corrección de contenido: ¿Hay errores conceptuales? ¿La respuesta correcta, lo es?
- Revisión de vocabulario apropiado: ¿El ítem tiene vocabulario apropiado?
- Revisión de sesgo: ¿El ítem ofende o penaliza?
- Revisión de errores de escritura: ¿El ítem presenta fallas de redacción de fácil corrección? (INEE, 2005b: 22).

Una vez que los integrantes del Comité analizaron el reactivo y respondieron las preguntas anteriores, se determinó que, si el dictamen de dos revisores era favorable, podía aprobarse el reactivo. Cabe recalcar que estos criterios guiaron todo el proceso de revisión de reactivos y que cada ítem estuvo sometido a la revisión de dos integrantes del Comité.

Luego de la revisión de los reactivos, en las sesiones teórico-prácticas, los equipos intercambiaron los ítems elaborados previamente, verificaron el cumplimiento de los puntos señalados en la Lista de cotejo y los comentaron en plenaria grupal. Una vez terminado este ejercicio, y ya familiarizados con el procedimiento, los integrantes del Comité proedieron a la elaboración de dos reactivos por equipo, el intercambio de los mismos para ser revisados con control de cambios y la Lista de cotejo, la corrección de los reactivos por sus elaboradores, atendiendo las observaciones hechas a los ítems. Esta dinámica se repitió una vez más, salvo que en esta última ocasión el trabajo fue a nivel individual y se terminó con una nueva plenaria en donde los integrantes del Comité expusieron sus comentarios acerca del proceso.

3.2. Ejemplos de reactivo

A continuación se ofrece un ejemplo de reactivo por cada uno de los niveles de logro académico que fueron determinados por un cuerpo colegiado de expertos en cada área disciplinar (Comité de Niveles de Logro y Puntos de Corte). Para PLANEA EMS 2017, se determinaron cuatro niveles en línea de ascenso: Nivel I: logro insuficiente de los aprendizajes. Nivel II: logro apenas indispensable de los aprendizajes. Nivel III: logro satisfactorio de los aprendizajes. Nivel IV: logro sobresaliente de los aprendizajes. Cabe resaltar que estos niveles son acumulativos, por lo tanto, los estudiantes que se encuentran en el Nivel IV demostraron suficiencia para responder correctamente los reactivos de los niveles anteriores.

- **Reactivo de Lenguaje y Comunicación para el Nivel I**

Lee la siguiente información y atiende lo que se solicita.

¿Conciertos para reducir el estrés?

Public Health

Por primera vez, investigadores han mostrado que ver música en vivo puede ayudar a la reducción de las hormonas del estrés en las personas incluyendo el esteroide, cortisol.

Durante la última década, un rango de estudios ha mostrado que escuchar música –incluso cantar– puede tener un impacto positivo en nuestro bienestar. Pero hasta ahora, nadie había buscado los beneficios de asistir a un concierto.

El estudio, del Colegio Imperial de Londres, muestra por primera vez que asistir a un evento cultural puede tener un impacto y disminuir el estrés (...).

Se midieron los niveles de las hormonas del estrés de los voluntarios antes de cada espectáculo y 60 minutos después del concierto. Entonces para evitar las variables, ambos conciertos tuvieron exactamente la misma música y duración.

El equipo descubrió que, en la mayoría de los participantes, los niveles de las hormonas del estrés –el cortisol y la cortisona– habían disminuido tras ambos conciertos.

El cortisol se produce cuando el cuerpo está bajo estrés tanto físico como mental y, aunque puede tener efectos positivos en pequeñas dosis (...) a largo plazo se ha asociado con un mayor riesgo de padecimientos cardiacos, diabetes e impotencia. Así que mantener esos niveles bajos es importante.

Uno de los hallazgos más interesantes es que no importaba qué tan musicales fueran los voluntarios –el concierto pareció tener un efecto similar en todos.

Aunque, aún falta más investigación respecto al tema, es un primer paso muy prometedor, por lo que nos tomaremos la libertad de recomendarte asistir a una tocada este fin de semana.

¿Conciertos para reducir el estrés? (Junio 2016). Muy Interesante.
Recuperado de: http://www.muyinteresante.com.mx/ciencia/16/04/14/musica-vivo-reduce-estres.html

Mario quiere hacer un mapa conceptual con los conceptos más importantes del texto, ¿cuál es la frase con la que debe iniciar su esquema?

A) Efectos de la música

B) Aumento del estrés

C) Conciertos de 60 minutos

D) La importancia de las investigaciones

Este reactivo está diseñado para saber si los estudiantes cuentan con la habilidad necesaria para identificar el tema central de un texto expositivo como el artículo de divulgación. En este ítem la mayor parte del contenido está enunciado de manera explícita y para resolverlo el alumno necesita, entre otros elementos, identificar la tipología textual a la que pertenece la lectura, reconocer cuál es el referente a lo largo de la lectura y determinar las palabras clave que le permitan identificar el tema central.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Texto expositivo	1	35728	81.7	0.29	81.7	7.45	3.1	7.65

- **Reactivo de Lenguaje y Comunicación para el Nivel II**

Relaciona el subgénero literario con la intención comunicativa que lo caracteriza.

Subgénero literario	Intención comunicativa
1. Poesía	a. Narrar hazañas
2. Fábula	b. Enseñar una lección
3. Epopeya	c. Producir emociones

A) 1c, 2b, 3a

B) 1c, 2a, 3b

C) 1b, 2c, 3a

D) 1b, 2a, 3c

Este reactivo se centra en evaluar si los estudiantes poseen la habilidad para identificar la intención comunicativa de algunos textos literarios como la poesía, la fábula y la epopeya. Para resolverlo, el alumno requiere reconocer cuáles son las características de cada uno de los géneros literarios señalados, identificar la función de la lengua que prevalece en ellos y conocer qué es la intención comunicativa para, finalmente, poder realizar la relación entre columnas.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Texto literario	1	35728	68.75	0.36	68.75	23.28	3.65	4.22

- **Reactivo de Lenguaje y Comunicación para el Nivel III**

La definición anterior hace referencia a:

A) los enunciados

Lee la siguiente información y atiende lo que se solicita.

Son elementos lingüísticos que contribuyen a dar orden y estructura al texto oral o escrito. Sirven para establecer relaciones tanto de forma como de significado entre las oraciones, proposiciones y párrafos. De esta manera se logra establecer una conexión clara entre los distintos fragmentos que componen el texto. Pueden ser adverbios, frases adverbiales o conjunciones.

La habilidad que los estudiantes deben demostrar para responder adecuadamente este reactivo se basa en identificar las características de los conectores discursivos (adverbios, frases adverbiales o conjunciones) que se emplean para dar coherencia y cohesión a un texto. Por ende, para contestar al ítem, el alumno debe movilizar su conocimiento sobre las características de un enunciado, de las conjunciones, de los modificadores y de los conectores; también debe identificar las diferencias entre oración, proposición y párrafo, reconocer el concepto de orden, de estructura y de elemento lingüístico y activar sus conocimientos previos sobre los mecanismos de cohesión y coherencia durante el proceso de redacción.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Texto argumentativo	4	35762	57.34	0.45	22.59	15.59	4.26	57.34

- **Reactivo de Lenguaje y Comunicación para el Nivel IV**

Lee la siguiente información y atiende lo que se solicita.

El uso de internet en los adolescentes

1. Internet se ha convertido hoy día en una herramienta indispensable en la vida de las personas. Sería difícil, especialmente para los más jóvenes, concebir un mundo en el cual "no estemos conectados".
2. Ingo Lackerbauer, en su libro "Internet", señala que la importancia de internet en el futuro desborda todo lo acontecido hasta ahora, se está convirtiendo en el "medio de comunicación global".
3. No hace falta explicar con detalles los beneficios de este maravilloso invento tecnológico. Nos permite educarnos, conocer, disfrutar. Es decir, es una herramienta multiuso. Precisamente, es este uso el que puede volverse negativo.
4. Estamos hablando de la adicción al internet. Muchos jóvenes pasan una gran parte del día navegando por páginas, publicando en las redes sociales o viendo videos en youtube. Usar el internet para el entretenimiento no es algo malo en sí. Lo malo es abusar. El mundo de la web está plagado de conocimientos muy útiles, lo ideal sería también utilizarse en esa faceta, y que no sea sólo como manera de ocio.
5. ¿Cuáles son los perjuicios que puede acarrear la adicción a internet? Debido a que el adolescente pasa un tiempo considerable frente al ordenador, una de las mayores consecuencias es la pérdida de una vida social activa. Es probable que pierda el contacto que tenga con sus amigos más cercanos, y pase más tiempo con los amigos "virtuales".

El uso de internet en los adolescentes. (Mayo 2016).
Recuperado de: http://www.milejemplos.com/lenguaje/ejemplos-de-textos-argumentativos.html

El enunciado que sustenta la postura del autor es:

A) Internet se ha convertido en una herramienta multiusos indispensable y favorable para la vida de hoy en día.

B) existen medios de comunicación global que se han vuelto un maravilloso invento tecnológico en la actualidad.

C) la adhesión del adolescente al mundo digital aparta a los jóvenes de disfrutar su entorno social inmediato.

D) el mundo de la web no sólo se utiliza para el ocio, sino para adquirir conocimientos muy útiles para todos.

El propósito de este reactivo es evaluar la habilidad de los estudiantes para identificar la postura del autor en un texto argumentativo como un artículo de opinión. Para contestar adecuadamente al ítem, el alumno debe emplear, entre otros recursos, su conocimiento sobre las características de un texto argumentativo, como el concepto de postura de un autor, identificar los argumentos que funcionan como fundamento de su opinión y la distinción entre los enunciados que expresan una idea subjetiva (opinión) y los que corresponden a información objetiva (hechos y datos). A nivel global, el estudiante debe ser capaz de identificar las palabras clave y comprender el texto de manera integral, ya que una parte importante del contenido se presenta de manera implícita, por lo que el alumno debe inferir la forma en que el autor organiza la información y la relación entre las diferentes partes del texto, que denotan la coherencia (unidad temática) del mismo.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Texto argumentativo	3	35728	32.17	0.22	33.45	7.01	32.17	27.23

- **Reactivo de Matemáticas para el Nivel I**

14. En la tabla se muestran los litros de agua purificada y sus precios correspondientes:

Litros de agua	12	14	15	18	25
Precio	$18.00	$21.00	$22.50		$37.50

Encuentra el dato que completa la tabla.

A) $30.50

B) $27.00

C) $25.50

D) $12.00

El estudio de la proporcionalidad es uno de los primeros acercamientos de los estudiantes con las relaciones entre variables, temática que se consolida con el estudio de las funciones. Se trata de un importante contenido de estudio en el nivel medio superior, por lo que este reactivo evalúa la habilidad del estudiante para reconocer el patrón de comportamiento proporcional que se establece entre los datos de la tabla y realizar los cálculos necesarios para determinar el valor faltante. Para resolver el ítem el alumno necesita realizar un procedimiento mediante división del precio a pagar entre los litros de agua y reconocer que en todos los casos el resultado es igual a 1.5. Esto significa que los datos son proporcionales y que un litro de agua tiene un precio de $1.50. Para obtener el valor faltante correspondiente al precio de 18 litros de agua, este valor deberá multiplicarse por 18, que es el dato que se tiene en la tabla.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Cambios y relaciones	2	35762	64.25	0.37	10.4	64.3	22.7	2.38

- **Reactivo de Matemáticas para el Nivel II**

6. Natalia compró cierta cantidad de chocolates que cuestan $8.00 cada uno. Al pagar con un billete de $50.00 recibió $18.00 de cambio, ¿cuál es la ecuación que permite obtener la cantidad de chocolates que compró Natalia?

A) $18x + 8 = 50$

B) $8x - 18 = 50$

C) $18x - 8 = 50$

D) $8x + 18 = 50$

La transición del pensamiento aritmético al pensamiento algebraico es un aspecto básico en la formación en la EMS, pues sienta bases importantes para la comprensión de diferentes fenómenos, desde el planteamiento de ecuaciones que representan una situación, como en el caso de este reactivo, hasta el planteamiento de funciones que describen el comportamiento de las variables de un fenómeno. Por lo tanto, este ítem se centra en evaluar la habilidad del estudiante para expresar algebraicamente la situación que se describe. Para resolverlo, el alumno deber ser capaz de realizar una serie de pasos: primero, representar el valor desconocido con una variable, en este caso con la x; segundo, comprender cada parte del problema y la relación de los datos con la variable, y tercero, expresar estas relaciones en forma algebraica.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Sentido numérico y pensamiento algebraico	4	35762	51.06	0.43	9.74	28.3	10.7	51.1

- **Reactivo de Matemáticas para el Nivel III**

18. Dadas las funciones $f(x) = x^3 - x^2 + 10$, $g(x) = 3x^2 + 2x - 10$ y $h(x) = 2x^3 + 2x^2 - 3x$, halla la función suma $f(x) + g(x) + h(x)$.

A) $3x^3 + 4x^2 - x$

B) $3x^3 + 6x^2 - x$

C) $6x^3 + 3x^2 - x$

D) $6x^3 + x^2 - x$

La comprensión de la noción de función, su simbolismo asociado y su manipulación algebraica (incluida la gráfica y tabular), es uno de los contenidos principales en el nivel medio superior, debido a que muchos fenómenos de diferentes disciplinas pueden expresarse mediante modelos matemáticos. La comprensión de su notación implica que el estudiante deje de ver la x como una incógnita, para comprenderla como una variable que puede tomar distintos valores. Además, la manipulación de las expresiones o representaciones de las funciones permite abstraer información implícita sobre el comportamiento y variación de las mismas. Por ende, el propósito de este reactivo es evaluar la habilidad de los alumnos para comprender la notación y operaciones con funciones. Para resolverlo es necesario que el estudiante active sus conocimientos sobre la notación de función y las reglas para sumar expresiones algebraicas polinómicas.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Cambios y relaciones	1	35762	36.3	0.36	36.3	24.1	25.5	13.1

- **Reactivo de Matemáticas para el Nivel IV**

8. ¿Cuál es el resultado de la siguiente división?

$$\left(-m^5n^2-\frac{1}{2}m^4n^4+\frac{2}{3}m^3n-4mn^4\right)\div\left(-4m^5n^3\right)$$

A) $\frac{1}{4n}+\frac{n}{8m}-\frac{1}{6m^2n^2}+\frac{n}{m^4}$

B) $\frac{1}{4n}-\frac{n}{8m}-\frac{1}{6m^2n^2}-\frac{n}{m^4}$

C) $\frac{1}{4n}+\frac{mn}{8}-\frac{1}{6m^2n^2}+\frac{n}{m^4}$

D) $\frac{1}{4n}+\frac{2n}{m}-\frac{8}{3m^2n^2}+\frac{n}{m^4}$

Resolver divisiones con polinomios y el manejo de las reglas de operación asociadas son habilidades significativas dentro el aprendizaje de las matemáticas, pues constituyen bases fundamentales para la manipulación algebraica de expresiones más complejas que combinan diferentes tipos de expresiones, como las expresiones algebraicas racionales. También sientan bases importantes para el dominio de reglas de operación más complejas como las asociadas con la derivación y la integración. Por ello, este reactivo se centró en evaluar la habilidad de los estudiantes para resolver divisiones con polinomios y el manejo de las reglas de operación asociadas. Se requiere un importante dominio y manejo del álgebra, por ende, para resolverlo, el estudiante movilizó, entre otros contenidos, su conocimiento de las reglas y leyes de los exponentes y los signos, así como la división de enteros y fracciones. El manejo coordinado de estos contenidos confirió a este reactivo una importante dificultad, al exigir un gran dominio de la operatividad algebraica.

Resultados del reactivo:

Área temática a la que pertenece el reactivo	RC	Alumnos	P por opción (%)					
			P%	Ptbis	1	2	3	4
Sentido numérico y pensamiento algebraico	1	35762	28.59	0.25	28.6	23.6	30.4	16.4

3.3. Edición del reactivo

La etapa de edición de los reactivos de opción múltiple fue el paso intermedio entre la entrega de los ítems por parte del Comité de Elaboración de Reactivos y el momento en que el Comité de Validez y Sesgo, entró en función para realizar el correspondiente. Durante este proceso, se estandarizó cada uno de los ítems a partir de los lineamientos establecidos por el INEE, los cuales buscan, ante todo, homogeneizar los criterios técnicos y el estilo de los reactivos. Esta fase fue un paso que apoyó el trabajo del Comité de Validez y Sesgo, ante todo, al estandarizar tanto el formato y el estilo, como algunos elementos técnicos de los reactivos a evaluar.

3.3.1. Revisión externa del reactivo

La Dirección General de Medición y Tratamiento de Datos (DGMTD) del INEE realizó la revisión externa. De acuerdo con el estatuto orgánico del INEE, esta Dirección tiene, entre otras funciones, la de: "Realizar la edición de instrumentos que requieran las mediciones y evaluaciones de componentes, procesos y resultados del Sistema Educativo Nacional, con la participación de las unidades administrativas que correspondan" (DOF, 31 enero 2017). La intervención de esta Dirección abarca desde la revisión técnica de los reactivos y la revisión de estilo hasta el diseño y edición de los reactivos, en congruencia con los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, del INEE, los cuales argumentan que es necesario:

> Revisar técnicamente los reactivos o tareas evaluativas antes de ser validados. Personal del organismo evaluador debe analizar los reactivos o las tareas para garantizar que éstos cumplan con los lineamientos técnicos que guiaron su elaboración. Si como resultado de la revisión técnica se identifica que es necesario algún ajuste, se debe recurrir al elaborador para realizarlo, también es posible presentarle propuestas y solicitar su apoyo cuando se requiera alguna precisión en el contenido del reactivo o la tarea evaluativa (INEE, 2017: 24).

Del mismo modo, es necesario:

> [...] verificar que el reactivo o la tarea no tenga errores ortográficos ni de redacción, con el fin de asegurar el uso adecuado del lenguaje y garantizar una comunicación eficaz. Es preciso definir criterios editoriales en un manual que permitan dar unidad editorial a los instrumentos de evaluación; estos criterios deben ser conocidos por las personas que participan en la construcción de los cuestionarios y pruebas y su revisión tales como elaboradores de reactivos y coordinadores de los instrumentos de evaluación, entre otros (INEE, 2017: 24).

Así, el trabajo de la DGMTD se centró en la identificar las áreas de mejora de cada uno de los ítems presentados por los miembros del Comité de Elaboración de Reactivos, a partir de detectar conceptos erróneos, contexto inapropiado, redacción inadecuada, error en la respuesta correcta, información innecesaria o ayuda no deseada, diseño gráfico inadecuado, entre otros elementos.

3.4. Comité de Validez y Sesgo

El cuerpo colegiado de este Comité revisó los aspectos lingüísticos y culturales de los reactivos, así como de sus características técnicas; verificó que se apegaran a las especificaciones base, carecieran de errores conceptuales, tuvieran una construcción técnicamente adecuada a los criterios establecidos por el INEE y no tuvieran sesgo cultural y lingüístico. Este procedimiento se justifica técnicamente porque "tiene como finalidad sustentar que los reactivos o tareas se alinean al objeto de medida y son adecuados para obtener información que permita hacer inferencias acertadas, también es útil para minimizar la intervención de variables irrelevantes al constructo medido" (INEE, 2017: 25).

El Comité de Validez y Sesgo contó con la participación de profesores del nivel medio superior de las diferentes entidades de la república mexicana y del campo disciplinar específico (Lenguaje y Comunicación o Matemáticas), con la finalidad de tener una amplia representación de las regiones del país, de los subsistemas que conforman la EMS y de las diversas formas de conceptualizar las diferencias culturales en un contexto intercultural. En la tabla 6 se muestra la relación de las instituciones participantes.

Tabla 6. Instituciones participantes en el Comité de Validez y Sesgo

Estado de la República	Lenguaje	Matemáticas
REGIÓN NOROESTE		
Baja California	DGECyTM	
Baja California Sur		DGETA
Chihuahua	Telebachillerato	
Sinaloa		DGECyTM
Sonora	DGETI	DGETI
REGIÓN NORESTE		
Coahuila		DGETI
Durango	DGETI	
Nuevo León		CECyTE
San Luis Potosí	DGB	
Tamaulipas		CONALEP
REGIÓN CENTRO		
México	Telebachillerato	Telebachillerato
Guerrero	Telebachillerato	
Hidalgo		DGETA
Morelos	DGETA	
Puebla		CECyTE
Tlaxcala	DGB	
REGIÓN OCCIDENTE		
Aguascalientes		DGETA
Colima	CONALEP	
Guanajuato	Telebachillerato	Telebachillerato
Jalisco	CONALEP	
Michoacán		CONALEP
Nayarit	CECyTE	
Querétaro		CONALEP
Zacatecas	DGB	
REGIÓN SURESTE		
Campeche		CONALEP
Chiapas	CECyTE	
Oaxaca		CECyTE
Quintana Roo	DGECyTM	
Tabasco	COLBACH	COLBACH
Veracruz		DGECyTM

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

3.4.1. Procedimiento metodológico para el Comité de Validez y Sesgo

De acuerdo con la metodología propuesta por el INEE, la primera tarea para el desarrollo del Comité de Validez y Sesgo debe ser la capacitación, la cual tiene como propósito dotar a los integrantes de los insumos y los lineamientos técnicos necesarios para la adecuada ejecución de sus tareas. "Se debe revisar el propósito de la validación, el procedimiento a seguir y los dictámenes que podrán asignar, haciendo hincapié en la importancia del proceso y su repercusión en la calidad del instrumento" (INEE, 2017: 25). Por ende, los jefes de prueba y la subdirectora del área abrieron las funciones del Comité con la contextualización en torno al Plan Nacional para la Evaluación de los Aprendizajes, desde sus objetivos generales hasta el proceso de diseño y conceptualización de PLANEA EMS 2017, poniendo énfasis en la determinación de los contenidos a evaluar, la descripción de los indicadores de aprendizaje y las especificaciones en las que se sustentó la fase de elaboración de reactivos.

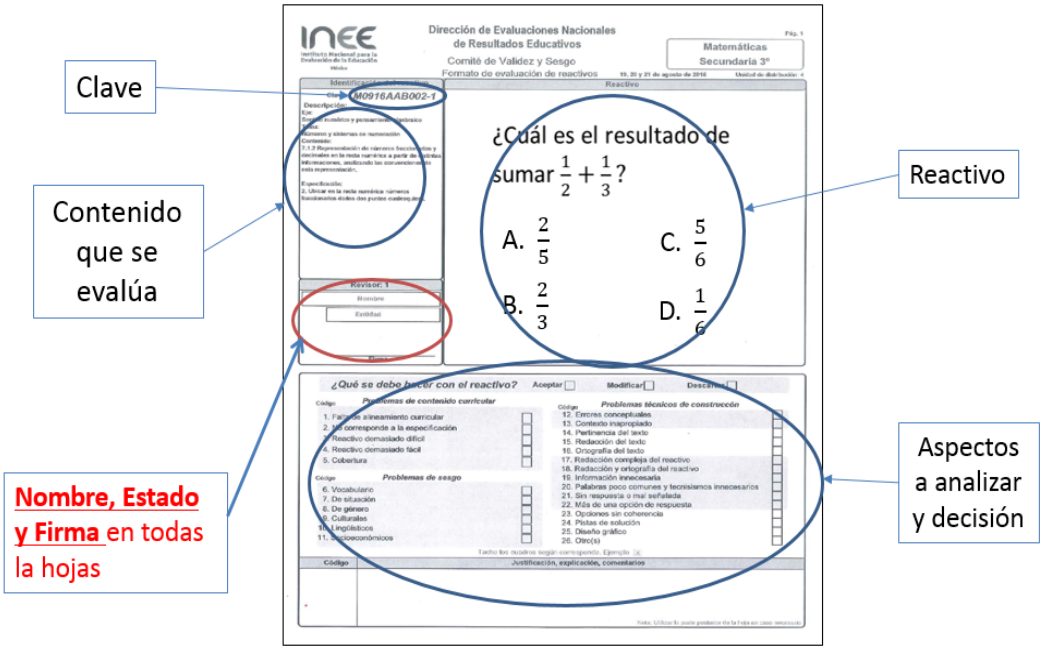
Posteriormente se le entregó a cada miembro del Comité la versión "corta" de los criterios de validación (ver Anexo V), en la cual se destacaron los aspectos centrales para la revisión y que, por ende, fue uno de los insumos necesarios para el desarrollo de las actividades del Comité. Finalmente, se procedió a la presentación de los ítems elaborados por el Comité de Elaboración de Reactivos, para que cada uno fuera revisado por dos expertos, con apoyo del personal del INEE. A lo largo de cada etapa de validación, los integrantes del cuerpo colegiado contaron con tres opciones de dictamen:

- **Aceptado.** El reactivo no requiere cambios.
- **Modificar.** Se proponen cambios para mejorar el reactivo.
- **Descartar.** El reactivo presenta problemas que no se pueden solucionar en el tiempo disponible, por lo que no puede pasar a la siguiente etapa.
- **Nueva valoración.** El reactivo se pospone para revisar al final.

Debido al carácter confidencial de los ítems, los integrantes del Comité de Validez y Sesgo sólo pudieron acceder al material de trabajo en el área y los tiempos asignados para este fin.

En el esquema 1 se muestra el formato de presentación de cada reactivo y los elementos que lo integraron durante la capacitación del Comité.

Esquema 1. Presentación de reactivos ante el Comité de Validez y Sesgo



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

El proceso metodológico de validación fue sistemático y estandarizado, de acuerdo con dos etapas de trabajo propuestas: individual y colegiada; asimismo, las personas que validaron dejaron evidencia de su aprobación en la última versión de los reactivos, como se detalla a continuación en el proceso que guió las actividades del Comité:

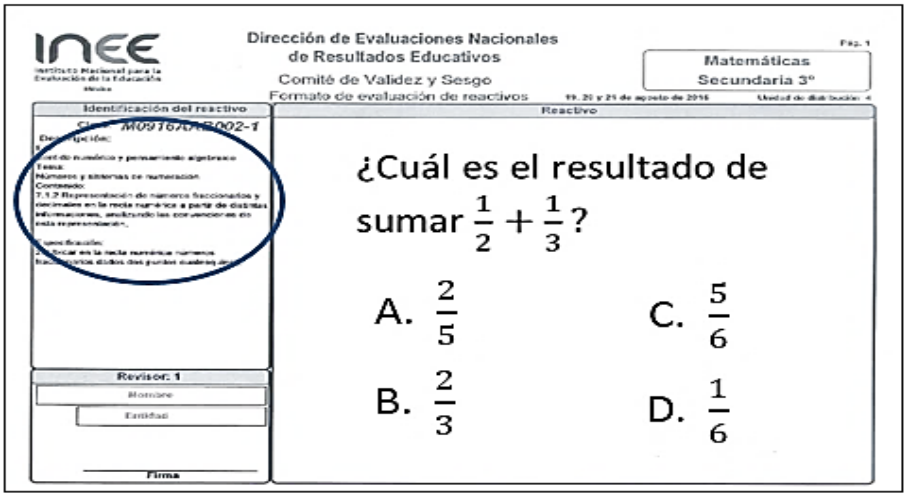
ETAPA 1.

TRABAJO INDIVIDUAL: los integrantes del Comité leyeron cuidadosamente el indicador de aprendizaje (especificación) del reactivo proyectado en una pantalla para este fin. Este proceso persiguió tres objetivos: que conocieran el contenido que el reactivo pretendía evaluar, que juzgaran su pertinencia y que validaran la alineación indicador de aprendizaje-reactivo.

TRABAJO COLEGIADO: a partir de este análisis, fue necesario que cada miembro presentara las evaluaciones realizadas, para determinar si efectivamente el indicador de aprendizaje tenía correspondencia con el reactivo, de esta manera: a) si los expertos coincidían en sus valoraciones, se consignaba el reactivo original, sin cambios; b) si los expertos no coincidían, se procedía a consensuar las valoraciones que se tuvieran.

Finalmente, los expertos procedieron a determinar el dictamen a partir de los criterios mencionados: Aceptado, Modificar, Descartar o Nueva valoración.

Esquema 2. Contenido del reactivo (etapa 1)



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

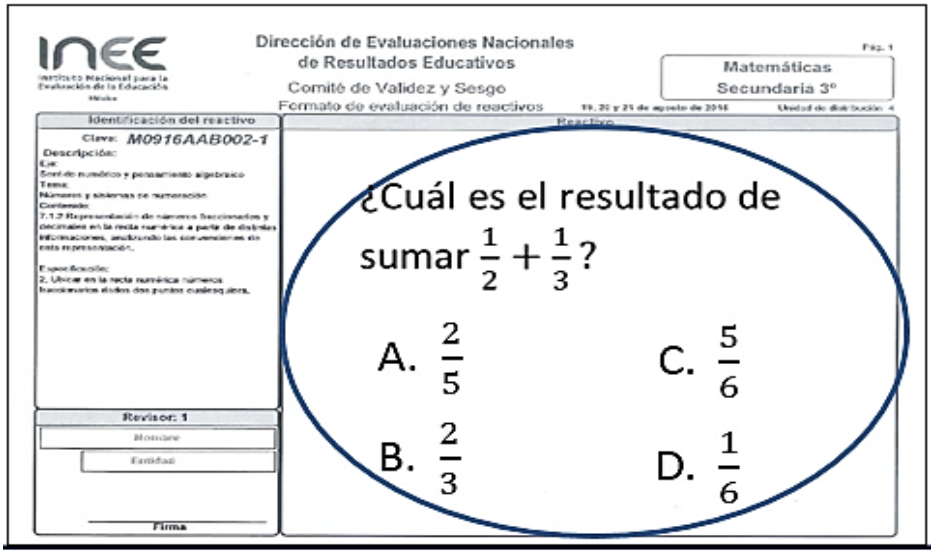
ETAPA 2.

TRABAJO INDIVIDUAL: el siguiente paso metodológico consistió en que los expertos de cada área disciplinar leyeran el reactivo y lo respondieran, ya que mediante este ejercicio, ellos podían conocer las habilidades y conocimientos necesarios para responder correctamente el reactivo, valorar la competencia escolar que se evaluaba, identificar la dificultad de la pregunta y valorar la pertinencia para la asignatura y el grado.

TRABAJO COLEGIADO: una vez respondidos los reactivos, los participantes del Comité nuevamente presentaron sus evaluaciones: a) si los expertos coincidían en sus valoraciones, se consignaba el reactivo original, sin cambios; b) si los expertos no coincidían, se procedía a consensuar las valoraciones que se tuvieran.

Finalmente, los expertos procedieron a determinar el dictamen a partir de los criterios mencionados: Aceptado, Modificar, Descartar o Nueva valoración.

Esquema 3. Responder el reactivo (etapa 2)



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

ETAPA 3.

TRABAJO INDIVIDUAL: esta labor fue la parte medular del proceso de validación de los ítems. El Comité de Validación y Sesgo debió valorar el reactivo conforme a los criterios de validez e identificar problemas o errores que devinieran en posibles fuentes de invalidez. Los aspectos que se tomaron en cuenta en esta etapa fueron:

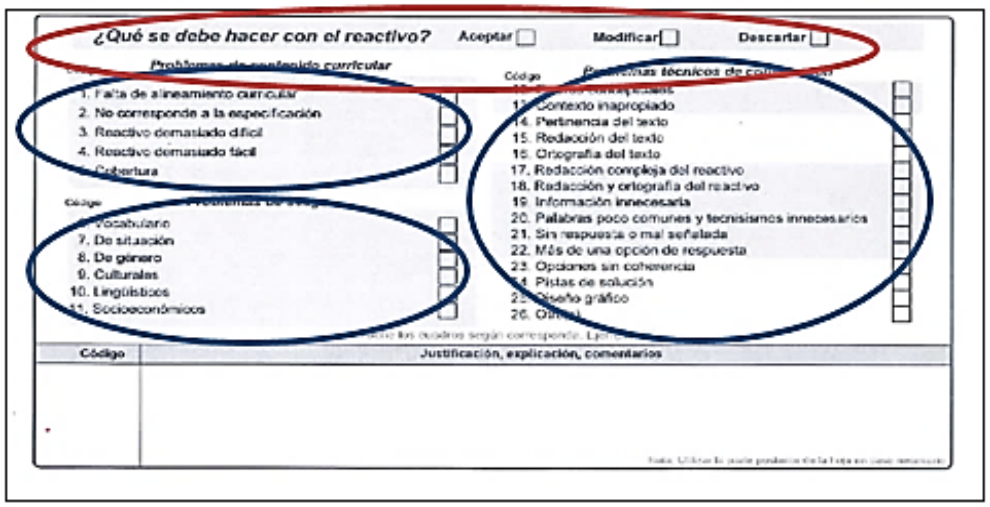
a) El contenido temático, para observar el grado en que los reactivos del examen representaban el dominio curricular a evaluar y la adecuación del nivel de dificultad con el grado escolar a evaluar.
b) La inclusión, para detectar los posibles sesgos social, cultural o de género en el vocabulario utilizado, las situaciones planteadas en la base del reactivo y los aspectos relacionados con estereotipos.
c) La construcción de los reactivos, para localizar conceptos erróneos, errores en la respuesta correcta, ayudas no deseadas, pertinencia de los elementos auxiliares, entre otros.

Como se mencionó anteriormente, en el Anexo V se puede consultar la versión "corta" de los criterios de validación establecidos por el INEE que fueron la guía para la revisión de los reactivos en esta etapa del procedimiento.

TRABAJO COLEGIADO: a partir de la detallada observación de estos elementos, los participantes del Comité nuevamente presentaron sus evaluaciones: a) si los expertos coincidían en sus valoraciones, se consignaba el reactivo original, sin cambios; b) si los expertos no coincidían, se procedía a consensuar las valoraciones que se tuvieran.

Finalmente, los expertos hicieron su dictamen a partir de los criterios anteriormente mencionados: Aceptado, Modificar, Descartar o Nueva valoración.

Esquema 4. Valorar el reactivo (etapa 3)



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

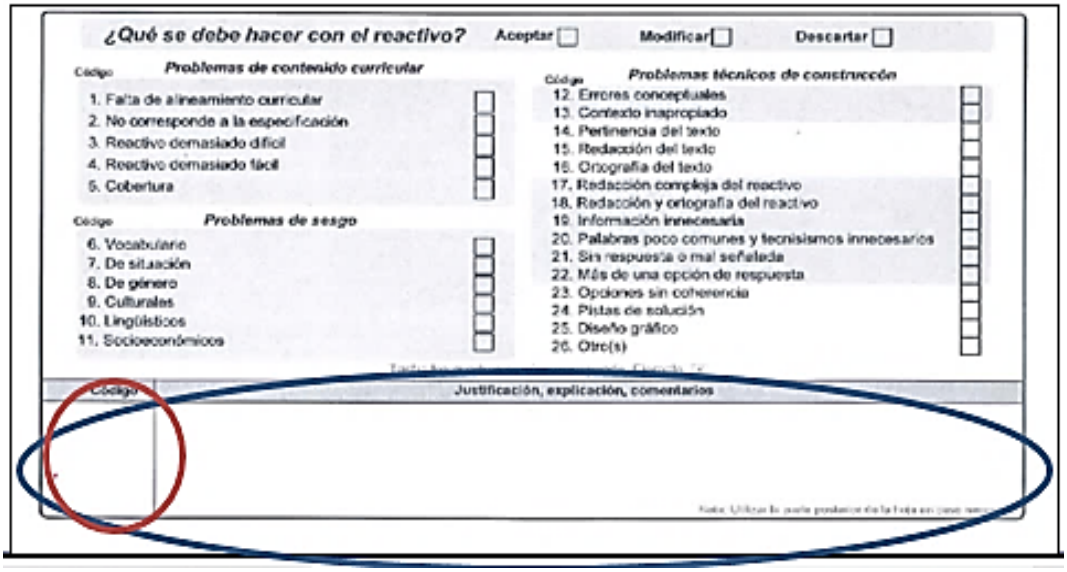
ETAPA 4.

TRABAJO INDIVIDUAL: los integrantes del Comité también tuvieron que justificar los errores o problemas identificados en el reactivo, con la finalidad de documentar el problema y las sugerencias de fondo y forma para corregir el reactivo, en caso de requerirlo.

TRABAJO COLEGIADO: una vez más se volvió a repetir el proceso del trabajo en conjunto, es decir, se presentaron las evaluaciones realizadas por cada miembro: a) si los expertos coincidían en sus valoraciones, se consignaba el reactivo original, sin cambios; b) si los expertos no coincidían, se procedía a consensuar las valoraciones que se tuvieran.

Finalmente, se realizó el dictamen a partir de los criterios anteriormente mencionados: Aceptado, Modificar, Descartar o Nueva valoración.

Esquema 5. Justificar los errores (etapa 4)



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Cuando fue necesario, las señalizaciones, correcciones o ajustes al reactivo se realizaron en el momento y sobre el cuerpo del mismo, para que fuera el Comité de Validez y Sesgo quien avalara la última versión.

Al término de las jornadas de trabajo, los coordinadores académicos de la prueba resguardaron en formato digital los reactivos elaborados, revisados y validados, como evidencia de quién los elaboró, los revisó y los validó.

Sin embargo, el proceso de revisión no terminó con el trabajo colegiado del Comité de Validez y Sesgo. Los reactivos fueron sometidos a tres procedimientos más, con la finalidad de afinar detalles para el piloteo de los mismos, los cuales fueron: revisión de estilo a los reactivos que sufrieron alguna modificación, ajuste en la edición de los reactivos que así lo requirieron y revisión y visto bueno de los reactivos modificados, procedimientos que se detallan a continuación:

- **Revisión de estilo a los reactivos que sufrieron alguna modificación:**

El proceso de verificación del reactivo, en cuanto a los errores ortográficos o de redacción, estuvo centrado ante todo en la comunicación eficaz y el uso adecuado del lenguaje que cada reactivo debió cuidar. Para llevar a cabo esta tarea, el INEE elaboró las *Pautas editoriales para la construcción de reactivos* (2016) que: "cuenta con normas gramaticales básicas que contribuyen al continuo empeño de mejorar la calidad de la redacción de los reactivos y, en consecuencia, facilitarle al sustentante la comprensión de éstos" (INEE, 2016c: 6).

En esta etapa, se revisaron los reactivos que fueron aceptados durante el Comité de Validación y Sesgo, a partir de los criterios que determina el manual de *Pautas editoriales para la construcción de reactivos*, los cuales comprenden los siguientes elementos:

Cuestiones gramaticales:

- Precisión y claridad: revisión del uso adecuado de conectores, tiempo verbal, adverbios, conjunciones, demostrativos, preposiciones y prefijos.
- Concordancia: correspondencia adecuada en la concordancia temporal, entre dos sustantivos, entre sujeto y el verbo, y entre sustantivo y adjetivo.
- Pluralidad de significados: prestar atención a la polisemia y el pleonasmo.
- Concisión con el léxico: observar tecnicismos, exceso de palabras y de pronombres, además del uso del queísmo y dequeísmo.
- Uso correcto de expresiones frecuentes (a donde - adónde - adonde / abajo - debajo de / acuerdo (de acuerdo con - de acuerdo a) / afuera - fuera de – afueras, entre otras).

Estilo editorial:

- Puntuación: uso correcto de punto y seguido, punto y coma, punto y aparte, punto final, dos puntos, puntos suspensivos, coma, comillas, paréntesis, guiones largos, guiones cortos, corchetes, signos de admiración e interrogación y acento (diacrítico).
- Pautas editoriales: uso de mayúsculas, minúsculas, cursivas (itálicas), negritas, extranjerismos, abreviaturas, números, títulos y subtítulos, tablas y gráficas, citas y referencias bibliográficas.

Ortotipografía para la elaboración de reactivos:

- Base del reactivo: donde se detalla el uso del punto final, la coma, el uso de mayúsculas, negritas, cursivas y la tipografía textual.
- Opciones de respuesta: que abarca el uso del punto final, la coma y otros signos de puntuación.
- Acotaciones para las pruebas de matemáticas: consideraciones para las unidades de medida en problemas aditivos y en problemas de perímetro, las expresiones algebraicas, el uso de signos matemáticos y las gráficas cartesianas (INEE, 2016c).

Posteriormente, se realizó el ajuste de los reactivos que así lo requirieron para uniformar los criterios técnicos, el formato y el estilo de los reactivos. Una vez revisada esta modificación, los reactivos fueron enviados a edición, formación e impresión.

Cabe aclarar que el sistema que se utilizó como herramienta, tanto para presentar los reactivos ante el Comité de Validez y Sesgo, como para sistematizar las observaciones de cada uno de los miembros, fue el sistema INEEDITA, el cual es un desarrollo de la Dirección de Aplicaciones Tecnológicas para el Evaluación (DATE), que pertenece a la Dirección General de Medición y Tratamiento de Datos (DGMTD) de la Unidad de Evaluación del Sistema Educativo Nacional (UESEN).

3.5. Prueba piloto

Un instrumento de evaluación está integrado por reactivos que fueron aplicados durante una fase de piloteo y cuyos resultados estadísticos fueron verificados cuantitativamente. En esta fase de piloteo se aplican los ítems a un grupo representativo de la población objetivo, para determinar su comportamiento estadístico y sus cualidades. Este proceso también permite probar el protocolo de levantamiento de datos que se replica durante la aplicación definitiva de la prueba, para prever posibles errores, defectos u omisiones procedimentales (INEE, 2017).

El piloteo de una prueba forma parte del marco metodológico que la implementación de un instrumento de evaluación debe seguir para obtener la confiabilidad de las subescalas. A este respecto, se puede hablar de una adecuada verificación cuantitativa en la medida en que la aplicación repetida de un instrumento a una población en condiciones similares arroja mediciones consistentes y estables. Por ello, el piloteo de los reactivos que conformaron PLANEA EMS 2017 fue un paso necesario para proveer información cuantitativa sobre los reactivos y, así, realizar ajustes en el proceso del armado de las formas para la aplicación definitiva.

La Dirección General de Levantamiento y Procesamiento de Datos (DGLPD) del INEE, diseñó la logística de aplicación y levantamiento de datos pues tiene, entre otras, la función de:

> Establecer mecanismos y protocolos para los procesos de levantamiento y procesamiento de datos de las mediciones y evaluaciones de componentes, procesos y resultados del Sistema Educativo Nacional, así como supervisar su cumplimiento y diseñar y planear los operativos de campo que requieran las evaluaciones y estudios del Instituto (DOF, 31 enero 2017).

3.5.1. Preparativos e integración de la prueba piloto

Para garantizar la equivalencia de las formas del instrumento y que a partir de ellas se pudieran derivar interpretaciones válidas, el ensamble de las formas fue un paso que se realizó cuidando aspectos para reducir al mínimo la varianza irrelevante al constructo. Este ensamble lo llevaron a cabo los coordinadores de la prueba y, a diferencia del armado de la prueba definitiva, sólo se consideraron los criterios cualitativos, como la representación del currículo de acuerdo con los ejes temáticos y los indicadores de aprendizaje de cada área disciplinar, la habilidad cognitiva necesaria para responder el reactivo, entre otros.

Para PLANEA EMS 2017, el primer paso en el armado de los cuadernillos para la aplicación piloto fue la elección de los reactivos. En total se pilotearon 355 reactivos para Lenguaje y Comunicación y 353 para Matemáticas. Una vez realizada esta selección, el segundo paso fue la distribución de los ítems en bloques de reactivos, a partir de los cuales se determinaron siete diferentes cuadernillos de la prueba para cada área disciplinar con 50 reactivos aproximadamente, los cuales se distribuyeron en función de:

1. Observar que cada eje temático tuviera una representación equivalente en cuanto a número de reactivos.
2. Cuidar que no se repitieran ítems que hicieran referencia al mismo indicador de aprendizaje.
3. Prestar atención a que las habilidades cognitivas necesaria para responder el reactivo fueran similares en cada bloque de ítems.
4. En el caso de la prueba de Lenguaje y Comunicación, atender a que el número de lecturas fuera semejante en las tres formas y que su extensión fuera más o menos equivalente; además, observar que el número de reactivos independientes (aquellos que no estuvieran anclados a una lectura) también fuera semejante.

3.5.2. Revisión de la versión piloto de la prueba

Una vez terminado el proceso de la distribución de bloques, el armado de los cuadernillos pasó por una etapa en la cual se revisó su legibilidad, tanto en la presentación de los ítems como en las

lecturas y los gráficos que les servían de base, con la finalidad de que no existiera ninguna carga cognitiva originada a partir de este elemento. También se cuidó que los ítems no tuvieran errores de redacción ni ortográficos y se hizo una nueva revisión de estilo. Finalmente, se revisó el diseño gráfico y la formación de la versión piloto (de los reactivos, de los gráficos y de los textos), a fin de asegurar que éstos se presentaran completos y sin dificultades de legibilidad (INEE, 2017: 29).

3.5.3. Envío a imprenta para generar cuadernillos

Una vez realizado el ensamble de los cuadernillos que constituirían la versión final de la prueba, éstas, desde el área de diseño, se integraron en un documento que fue enviado a la imprenta. La selección de la imprenta se realizó bajo el más estricto protocolo de seguridad y confidencialidad para disminuir el riesgo de ser usada con fines diferentes a sus objetivos. Personal del Instituto, específicamente capacitado para esta tarea, y de la empresa responsable de imprimir la prueba se reunieron en una sesión de trabajo en la cual se revisó que la versión de la prueba que se imprimiría sería la misma que la que se entregó en el archivo electrónico final.

3.5.4. Revisión de los cuadernillos y autorización para su reproducción

Esta fase consistió en que el personal del INEE seleccionó un número de materiales al azar para revisar que correspondieran a la versión original de la prueba. La revisión estuvo a cargo de personal del INEE y cada miembro de este grupo revisó el mismo material seleccionado. Para ello fue indispensable: "contar con las pruebas de impresión firmadas por los responsables, que constituye[ro]n la versión autorizada de contenido y forma, de cómo deber[ía]n recibirse los ejemplares solicitados" (DGLPD, 2017a: 5). Una vez que se realizó dicha verificación, entonces se autorizó, mediante carta de aceptación, la reproducción del número total solicitado para la prueba.

3.6. Aplicación de la prueba piloto

En el mes de diciembre de 2016, se llevó a cabo el piloteo de los instrumentos: las pruebas de logro de aprendizajes en Lenguaje y Comunicación y en Matemáticas.

Para elaborar los reportes estadísticos del comportamiento de cada reactivo, los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* determinan que: "el tamaño de muestra tiene que ser suficiente para obtener estadísticos estables y con la precisión suficiente que permita tomar decisiones bien informadas sobre la inclusión, ajuste y eliminación de tareas evaluativas o reactivos que integrarán los instrumentos" (INEE, 2017: 27). En este sentido, el objetivo del piloteo fue captar, como mínimo, 200 respuestas por ítem. Además, en los grupos en que se planeó realizar el análisis de comportamiento diferencial, se buscó una muestra de 300 respuestas para poder observar la tendencia del reactivo a funcionar de manera diferente en distintas subpoblaciones.

Así, la población objetivo se integró por centros escolares pertenecientes a los diversos subsistemas de la EMS y que se encontraban en diferentes entidades federativas, ya que esta integración permitió la realización de estudios comparativos, a fin de detectar posibles sesgos en los reactivos.

A continuación se detalla el procedimiento operativo que se siguió para dicha aplicación.

3.6.1. Determinación de la muestra para el piloteo

Como ya se mencionó, para determinar la población objetivo (grupo representativo), se tomaron como base los estándares mínimos recomendados para obtener resultados válidos de los ítems y de los análisis de comportamiento diferencial: 200 respuestas por cada reactivo y 300 para observar el comportamiento diferencial.

La muestra se definió a conveniencia, ya que cada entidad participante "fue responsable de elegir a las escuelas de acuerdo con la facilidad para llegar a ellas siempre y cuando satisficieran la cuota por estrato educativo impuesta en la distribución" (DGMTD, 2017: 3). Por cuestiones operativas, se tomó la determinación de designar un máximo de 35 alumnos por cada aplicador.

En la siguiente tabla se muestran los diferentes subsistemas que conformaron la población del piloteo de reactivos de PLANEA EMS 2017.

Tabla 7. Subsistemas que constituyeron la muestra de la prueba piloto

Bachillerato de las Universidades Autónomas (Autónomo)	Dirección General de Educación Tecnológica Agropecuaria (DGETA)
Bachilleratos estatales (Bachilleratos estatales)	Dirección General de Educación Tecnológica Industrial (DGETI)
Centro de Educación Científica y Tecnológica (CECYTE)	Educación Media Superior a Distancia (EMSAD)
Colegio de Bachilleres (Colegio de Bachilleres)	Otros Subsistemas (Otros Subsistemas)
Colegio Nacional de Educación Profesional Técnica (CONALEP)	Bachilleratos Particulares (Privado)
Tele Bachillerato Estatales (TeleBach Estatales)	

Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

3.6.2. Diseño de la aplicación piloto

El área encargada del levantamiento de datos fue la Dirección General de Levantamiento y Procesamiento de Datos (DGLPD), encargada de diseñar y realizar los operativos en campo, así como de establecer las normas técnicas de operación en cada levantamiento. El inicio formal del trabajo de campo en las entidades federativas contó con la participación del subdirector de Diseño y Operación, del director de Operación de Campo y de la misma DGLPD, tal como el INEE ha establecido en su *Manual de procedimiento para el inicio formal del trabajo de campo para el levantamiento y procesamiento de datos en las entidades federativas* (DGLPD, 2017b). En esta segunda etapa del procedimiento también colaboraron el responsable operativo de cada área estatal de evaluación programada en la muestra, el coordinador regional responsable del levantamiento de datos en su región, el coordinador-aplicador responsable de organizar la aplicación en la escuela, y el aplicador, quien administró en un grupo los instrumentos de evaluación.

El personal del INEE coordinó y supervisó el empaquetado de los materiales de aplicación empleados durante el levantamiento de datos de la prueba piloto, el primer paso metodológico para llevar a cabo el levantamiento de datos fue la notificación tanto a las Direcciones del INEE en las entidades federativas (DINEE) como a las autoridades educativas locales, mediante comunicado. Una vez concluido este paso, el siguiente fue el envío de la información inicial a las DINEE, con al menos los siguientes elementos: a) Muestra para el piloteo, b) Calendario de actividades, c) Descriptor de actividades.

- **Capacitación de los aplicadores de la prueba piloto**

Una de las actividades previas al levantamiento fue la capacitación de aplicadores externos; es decir, personal que no formaban parte de la comunidad escolar en la que desempeñaron sus funciones. Esto ha sido un lineamiento específico para la aplicación de las pruebas de evaluación del logro educativo que el Instituto ha delimitado con la finalidad de prevenir la inflación de los resultados de manera artificial.

Cada aplicador debió cumplir con un perfil específico: que fuera mayor de edad; que contara, como mínimo, con el bachillerato concluido; que tuviera experiencia en la aplicación de instrumentos de evaluación y el manejo de documentos diversos; que tuviera iniciativa para la solución de problemas, disponibilidad de tiempo completo y una actitud de respeto y responsabilidad.

La capacitación abarcó, principalmente, la exposición acerca de los propósitos de PLANEA EMS 2017, la explicación de la metodología de aplicación y del protocolo de seguridad del instrumento de evaluación, así como de las funciones del aplicador. De igual forma, se detallaron algunos mecanismos para la atención de dudas de los estudiantes y la resolución de algunos problemas (tanto técnicos como metodológicos) que pudieron presentarse el día de la administración.

- **Protocolo de levantamiento de datos para la aplicación piloto**

Durante la planeación, se definió que la administración de la prueba piloto se haría en dos días. A cada estudiante se le entregó un cuadernillo de cada prueba (Lenguaje y Comunicación y Matemáticas), con 50 reactivos de opción múltiple y una hoja de respuestas para lector óptico, con sus datos. La prueba se resolvió sólo con lápiz y borrador, no se permitió el uso de materiales de apoyo. El instrumento de evaluación fue autoadministrable, por lo que cada estudiante contestó de manera individual y de acuerdo con las instrucciones que se estipulaban en éste.

El diseño para el levantamiento de datos se realizó de la siguiente manera:

<u>Jornada 1:</u>

- El aplicador comenzó con la "organización de la aplicación", la cual contempló una duración aproximada de 1 hora. Esta organización incluyó el contacto inicial con el director o responsable del plantel y la preparación de la prueba de Lenguaje y Comunicación.
- Posteriormente, el aplicador inició la prueba de Lenguaje y Comunicación, con una duración aproximada de 150 minutos.
- Al cierre del día 1 se hizo un control de datos y del material, con una duración aproximada de 30 minutos.

<u>Jornada 2:</u>

- El aplicador inició la jornada con la "organización de la aplicación"; durante aproximada de 1 hora, preparó la aplicación de la prueba de Matemáticas.
- Inmediatamente después, entregó los cuadernillos de Matemáticas a los estudiantes. La prueba tuvo una duración de 120 minutos.

El levantamiento de datos se realizó durante los días hábiles escolares, por ello, se previó que las escuelas en las que se aplicaron los instrumentos de evaluación tuvieran un horario compatible con los tiempos establecidos en el protocolo de aplicación.

3.6.3. Protocolo de seguridad para la administración de la prueba piloto

Fue necesario garantizar la confidencialidad DE PLANEA EMS 2017, debido a que los reactivos piloteados, una vez realizado el análisis estadístico, conformaron el instrumento de evaluación definitivo. Por esta razón, las normas que se siguieron durante la aplicación fueron las siguientes:

- Se informó a la comunidad escolar sobre su participación en la administración de la prueba piloto.
- Sólo el aplicador y los alumnos preseleccionados para la muestra pudieron permanecer dentro del aula designada para la administración de la prueba.
- El aplicador debió centrarse en la guía expresamente diseñada para apoyarlo en su labor.
- Los alumnos que terminaron antes de lo previsto tuvieron que entregar todo su material al aplicador y dirigirse al aula asignada por el director.
- Todo el material que no fue utilizado durante la administración de la prueba piloto debió ser resguardado, en todo momento, por el aplicador.
- Una vez terminada la jornada, el aplicador mantuvo en custodia todos los materiales de la aplicación hasta recibir indicaciones del coordinador-aplicador.

- **Informes de la aplicación piloto**

Tras el levantamiento de datos, la DGLPD fue la encargada de documentar los procesos y los resultados obtenidos de éste, a partir de solicitar a las direcciones correspondientes la información sobre la organización, desarrollo, resultados y recursos utilizados en el levantamiento de datos. Con esta información pudo llevar a cabo la elaboración, revisión y autorización del Reporte Técnico, conforme al contenido temático definido para tal efecto. Este documento se almacenó, además, en un archivo electrónico con la finalidad de que constituyera parte del acervo histórico de la DGLPD.

3.6.4. Protocolo para el resguardo de la información

Otra de las pautas importantes con las que PLANEA EMS 2017 debió cumplir fue el resguardo de la información, con el fin de garantizar la seguridad y confidencialidad de los materiales y la información antes, durante y después de la administración del instrumento de evaluación. Este procedimiento incluyó las siguientes normas:

- El responsable operativo entregó los instrumentos de evaluación a cada coordinador-aplicador, en cajas selladas (una por cada área disciplinar). Éste debió verificar que ninguna etiqueta estuviera rota y, de lo contrario, reportarlo inmediatamente.
- El día de la administración del instrumento, el coordinador-aplicador trasladó los materiales al plantel correspondiente, donde el director realizó la misma verificación de los sellos de cada caja.
- Los materiales se entregaron al aplicador, quien nuevamente verificó los sellos y confirmó que la cantidad de instrumentos de evaluación correspondía con el número de alumnos del grupo a su cargo.
- Al final de la jornada, los materiales se contabilizaron en presencia del director y el coordinador-aplicador le entregó los cuadernillos utilizados y no utilizados en un paquete debidamente sellado.

- Las hojas de respuestas (utilizadas y no utilizadas) se empaquetaron, se etiquetaron y se sellaron. Los responsables firmaron de conformidad con el procedimiento.
- Como último paso, el coordinador-aplicador trasladó las cajas y las entregó al responsable operativo.

Los pasos anteriores fueron complementados con evidencias para la verificación documental. En el caso del resguardo de materiales, éste fue acompañado por un inventario en el que se asentó la cantidad de material que se tenía bajo resguardo, con la finalidad de llevar un estricto control de éste y facilitar su localización.

3.7. Estadísticos de los reactivos piloteados

Con la finalidad de estimar con mayor precisión el desempeño de los estudiantes evaluados a nivel nacional por PLANEA EMS 2017, el análisis psicométrico de los instrumentos que integraron la prueba fue un factor importante que se realizó previamente a la aplicación y emisión de resultados definitivos, ya que a partir de este análisis se pudo mejorar la calidad técnica del instrumento de la prueba nacional; esto para proporcionar resultados confiables, acordes con el objetivo de la evaluación. Para un mayor control, la información arrojada sobre estos reactivos se archivó en informes técnicos, que posteriormente apoyaron la elección de los ítems que conformarían la prueba definitiva mejorando la calidad métrica del instrumento.

Se aplicaron dos tipos de análisis: el psicométrico y el de comportamiento diferencial (DIF). El primero mostró la distribución del logro de los estudiantes evaluados, realizando una comparación entre las habilidades estimadas para responder el reactivo y los parámetros de dificultad de éstos. Este análisis: "se realizó utilizando la Teoría clásica del test. El software utilizado fue R, en específico el paquete CTT (*Classical Test Theory Functions*), y JMetrik. Se generaron resultados a nivel global y para los subsistemas" (DGMTD, 2017: 4).

Mientras que el segundo análisis se refiere a la distancia entre un valor y el cero, medida con la que se compararon subpoblaciones de estudiantes que, contando con el mismo nivel de habilidad para responder un reactivo, no tenían la misma posibilidad para hacerlo correctamente.

> El análisis de comportamiento diferencial de reactivos (*Differential Item Functioning*, DIF) se realizó utilizando el modelo Logístico de Rasch-TRI (Teoría de respuesta al ítem). El software que se utilizó fue R, en específico el paquete TAM (*Test Analysis Modules*). Los resultados del DIF se generaron únicamente a nivel Global debido a que se requiere un tamaño de muestra mínimo de respuestas por reactivo que solo se alcanza en ese nivel (DGMTD, 2017: 4).

En el caso de PLANEA EMS 2017, el DIF generó una subescala de acuerdo al género: hombres y mujeres, donde el grupo de referencia (a quien se espera que la prueba favorezca) fue el de los hombres y el grupo focal (a quien se espera que la prueba ponga en situación de desventaja) fue el de las mujeres.

Los estadísticos que se generaron se calcularon para cada reactivo a partir de los dos análisis mencionados y para cada subpoblación (hombres y mujeres). Así, se obtuvieron cinco índices, de acuerdo con el reporte de la DGMTD:

1. Dificultad del reactivo (P %): se define como el porcentaje de individuos en la muestra que respondieron correctamente al reactivo.
2. Índice de discriminación (ID): proporciona información sobre la capacidad de un reactivo para diferenciar entre los individuos examinados de alta y baja puntuación. Se calcula mediante

$$ID=P_{High}-P_{Low}$$

Donde:

P_{High}: es la proporción de examinados que respondió al reactivo correctamente y que se encuentra en el grupo de 27% de los sujetos de más alta puntuación en la prueba.

P_{Low}: es la proporción de examinados que respondió al reactivo correctamente y que se encuentra en el grupo de 27% de los sujetos de más baja puntuación en la prueba.

3. Correlación punto biserial corregida (Ptbis): coeficiente de correlación que indica el grado de relación entre las respuestas de un reactivo dicotómico y la puntuación en la prueba, sin incluir al reactivo en cuestión.
4. P por opción (%): porcentaje de individuos en la muestra que respondieron a cada una de las opciones de respuesta del reactivo.
5. Ptbis por opción: coeficiente de correlación punto biserial corregida de cada una de las opciones de respuesta del reactivo (DGMTD, 2017: 4).

A nivel global también se calcularon estadísticos, los cuales se concentraron en tres índices de resultados:

6. Dificultad: dificultad del modelo de Rasch en la escala de PLANEA.
7. Infit: ajuste interno del modelo de Rasch.
8. DIF: análisis Rasch-DIF.

3.7.1. Análisis de resultados de los reactivos de la prueba piloto

En atención a una mayor claridad del proceso, a continuación se presentan los análisis y resultados estadísticos de la prueba piloto, así como los criterios seguidos para la integración de la prueba definitiva, con base en los resultados estadísticos.

- **Calibración**

Tanto en la aplicación piloto como en la definitiva, una vez que se eliminaron los reactivos con grandes deficiencias estadísticas, fue necesario llevar a cabo la fase de calibración, la cual consistió en la estimación de los parámetros de dificultad de los reactivos para el modelo logístico simple de Rasch. Para la calibración de PLANEA EMS 2017, el modelo condicional de respuesta al ítem fue usado en conjunción con el modelo de la población, sin utilizar variables de condicionamiento.

Además, el método que se usó para realizar la estimación de los parámetros de los ítems fue el de máxima verosimilitud marginal. Este método consiste en asumir que la habilidad es un efecto aleatorio y que existe independencia condicional entre las respuestas. Así pues, hay posibilidad de suprimir los parámetros de las personas del proceso de estimación integrando sobre la habilidad; esto se reduce a maximizar el logaritmo de la verosimilitud que surge de la siguiente ecuación:

$$p(\boldsymbol{X}|\boldsymbol{\alpha},\boldsymbol{\delta}) = \prod_{i=1}^{N} \int p(\boldsymbol{x}_i|\theta,\boldsymbol{\delta}) p(\theta|\boldsymbol{\alpha})\, d\theta$$

La calibración realizada a cada una de las escalas de PLANEA EMS 2017 se hizo independientemente de los parámetros de las otras y, para ella, se utilizó toda la muestra de estudiantes a los que se les aplicó la prueba piloto.

- **Limpieza y validación de datos**

Antes de iniciar formalmente los análisis estadísticos de los reactivos piloteados, un paso metodológicamente previo fue el de la limpieza y validación de datos, el cual consistió en detectar errores e inconsistencias de los cuadernillos, en la corrección de datos (cuando fue posible), en la asignación de códigos de identificación, la determinación del tipo y ancho más adecuado a las variables y la integración de todo en una base de datos consistente, para que de esta manera éstos fueran confiables para los análisis posteriores.

En el caso de PLANEA EMS 2017, este proceso siguió una serie de pasos que se muestran en el Diagrama 1.

Diagrama 1. Proceso de limpieza y validación de datos	
1.	Cotejar la información recibida contra la solicitada en las especificaciones de lectura.
2.	Verificar que los nombres, tipos y anchos de todas las variables, formas e instrumentos, correspondan con lo esperado según las especificaciones de lectura.
3.	Verificar que todas las variables de todas las formas y cuestionarios se encuentren en los dominios de cada una de ellas, no debe haber valores no esperados.
4.	Calcular los porcentajes de respuesta omitida y respuesta múltiple para cada variable de cada forma de cada instrumento.
5.	Extraer una muestra aleatoria simple de los registros de las bases de datos y solicitar los cuadernillos correspondientes, para verificar que la información electrónica es fidedigna de la información física.
6.	Recodificar las variables según sea necesario para un manejo adecuado de datos.
7.	Generar archivos finales en el formato más óptimo para su análisis, incluyendo variables de otras fuentes de información.

Fuente: Dirección General de Medición y Tratamiento de Datos.

Una vez realizado dicho procedimiento, se notaron casos en los que no fue posible corregir los datos, así que éstos fueron identificados con códigos que permitieron una sencilla ubicación posterior.

Tabla 8. Códigos de errores		
Código	Descripción	Tipo de evaluación
94	No se presentó a la evaluación a la que pertenece este reactivo	Logro y contexto
97	Reactivo no administrado (No se incluyó en su cuadernillo de examen)	Logro
98	Respuesta múltiple	Logro y contexto
99	Respuesta omitida	Logro y contexto

Fuente: Dirección General de Medición y Tratamiento de Datos.

3.7.2. Análisis estadístico de los reactivos

El análisis estadístico de los reactivos es un procedimiento que consiste en identificar aquellos ítems que presenten alguno de los siguientes problemas:

1. Dificultad extrema: ocurre cuando el porcentaje de respuestas correctas es menor a 10% o mayor a 90%. En los reportes se marcan con una (*P*).

2. Un distractor tiene una correlación punto biserial corregida mayor a la de la respuesta correcta. Esto puede deberse a que la respuesta fue mal especificada, o que el reactivo tenga más de una respuesta correcta, o que el reactivo tenga problemas de construcción, que el contenido medido sea ajeno al dominio por la población evaluada, etc. En los reportes se marca con una (*K*).
3. La correlación punto biserial entre la respuesta del reactivo y el puntaje total es menor a 0.15. En el INEE los reactivos con una ptbis menor a 0.15 han sido excluidos de los análisis. Este criterio es bastante laxo en comparación con el utilizado para las pruebas PISA y NAEP de 0.20. En los reportes se marca con una (*B*).
4. La correlación punto biserial entre la respuesta del reactivo y el puntaje total es menor a 0.20. Como se mencionó arriba es el criterio utilizado por PISA y NAEP y, a futuro también a implementarse para las pruebas del INEE una vez que la calidad de las pruebas se incremente. En los reportes se marca con una (*b*).
   Para el modelo logístico de Rasch, Se indica si el reactivo se desajusta al modelo (infit). Dicho ajuste debería estar entre 0.8 y 1.2. En los reportes se denota con una (*I*) (DGMTD, 2017: 8).

- **Modelo de puntuación de las respuestas**

Para PLANEA EMS 2017 se utilizó como modelo de puntuación los Mapas de Wright (mapas ítem-person), empleando la escala 200-800. Esta decisión, de acuerdo con la DGMTD, estuvo basada en la facilidad de interpretación mediante dicha escala, ya que ésta ha sido utilizada en evaluaciones anteriores, como PLANEA 00, PLANEA 06 y PLANEA 09, "puesto que en cada forma la media de los reactivos es 500 (0 en escala Logit)" (DGMTD, 2017: 5). También se utilizó el Modelo de Rasch para detectar el DIF, el cual:

> [...] compara dos grupos mediante la estimación de la dificultad de los reactivos para cada uno de los grupos, i.e., se estiman $\delta i,$, $\delta i,F$, donde se considera al grupo focal (F) y al grupo de referencia (R). (Wright, B.D., & Stone, M.H., 1979), (Wright, B.D., Mead, R.J, & Draba, R., 1976) y (Wright, B.D., & Masters, G.N., 1982). La diferencia media en logro entre ambos grupos se controla fijando la dificultad promedio de los reactivos en cero: $\Sigma\delta i,=1= \Sigma\delta i,=1=0$.
>
> La diferencia en logits entre las dificultades de reactivos estimadas para ambos grupos es una medida para determinar el Funcionamiento diferencial de reactivos: $\Delta i,= \delta i,- \delta i,F$.
>
> Bajo la hipótesis de que no se presenta DIF, el valor esperado de Δi, es 0.
>
> Una vez identificados los reactivos con funcionamiento diferencial, se procede a clasificarlos en tres categorías, bajo el siguiente criterio (DGMTD, 2017: 6-7).

Tabla 9. Funcionamiento diferencial de reactivos (DIF)		
Categoría	**Criterio**	**Codificación**
Insignificante que favorece al grupo de referencia	$\Delta i,R<0$ y $\|\Delta i,R\|<0.426$	A-
Insignificante que favorece al grupo focal	$\Delta i,R>0$ y $\|\Delta i,R\|<0.426$	A+
Intermedio que favorece al grupo de referencia	$\Delta i,R<0$ y $0.426\leq\|\Delta i,R\|<0.638$	B-
Intermedio que favorece al grupo focal	$\Delta i,R>0$ y $0.426\leq\|\Delta i,R\|<0.638$	B+
Grande que favorece al grupo de referencia	$\Delta i,R<0$ y $\|\Delta i,R\|\geq0.638$	C-
Grande que favorece al grupo focal	$\Delta i,R>0$ y $\|\Delta i,R\|\geq0.638$	C+

Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

Como se mencionó anteriormente, los subgrupos se definieron por la variable de género, en donde el grupo de referencia fue el de los hombres y el grupo focal, el de las mujeres.

- **Reporte final del análisis estadístico de los reactivos**

A continuación se presentan los informes específicos de los reactivos piloteados en PLANEA EMS 2017, por asignatura (Lenguaje y Comunicación y Matemáticas), los cuales fueron presentados por la DGMTD en su informe titulado *Piloteo PLANEA 12, aplicado en 2016 (Resultados del análisis psicométrico de reactivos de logro de opción múltiple)*, informe con fecha del 5 de enero de 2017 y que contiene los análisis estadísticos que sirvieron para seleccionar los ítems que conformaron la prueba definitiva de PLANEA EMS 2017.

LENGUAJE Y COMUNICACIÓN

Tabla 10. Estadísticos clásicos de las formas de Lenguaje y Comunicación (análisis global)							
Indicador	**Forma 01**	**Forma 02**	**Forma 03**	**Forma 07**	**Forma 08**	**Forma 09**	**Forma 10**
N de reactivos	52	52	51	50	50	50	50
N de examinados	789	797	804	504	505	489	515
Mínimo	7	0	8	7	8	10	0
Máximo	36	39	41	34	37	39	38
Media	22.479	21.589	24.672	19.101	20.661	23.658	23.33
Mediana	22	22	25	19	20	23	23
Desviación estándar	5.319	5.309	6.307	4.645	5.066	5.385	5.485
Rango intercuantil	7	7	9	6	7	7	7
Sesgo	0.037	-0.009	-0.168	0.249	0.362	0.116	-0.078
Curtosis	-0.327	0.0001	-0.564	-0.098	-0.039	-0.302	0.205
KR21	0.5597	0.563	0.693	0.462	0.538	0.581	0.598
Coeficiente Alfa	0.618	0.619	0.721	0.553	0.597	0.631	0.663
SEM	3.2877	3.2731	3.333	3.106	3.216	3.273	3.183
Media P	43.229	41.519	48.378	38.202	41.323	47.317	46.660
Media correlación punto biserial	13.090	12.912	18.397	10.593	12.616	14.095	15.464

Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

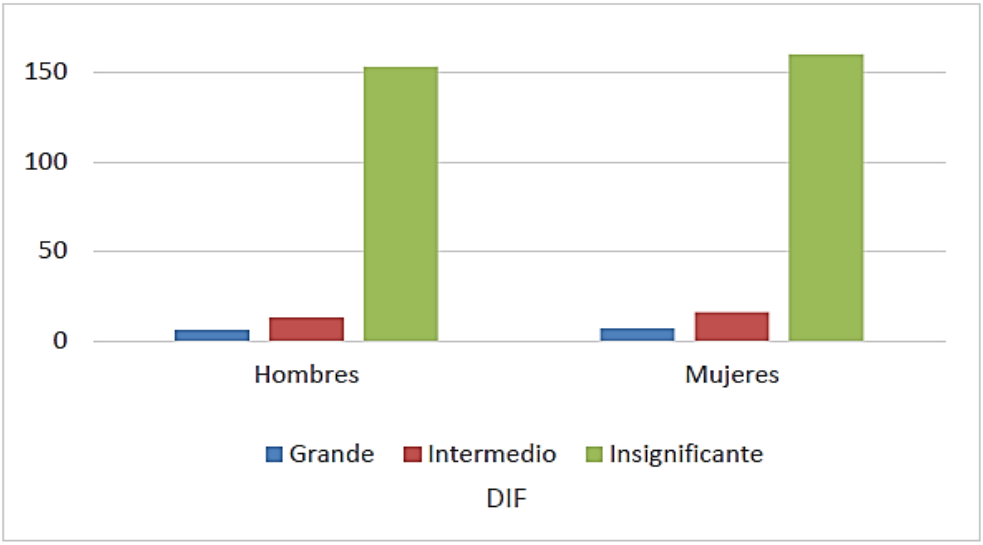
Tabla 11. Lenguaje y Comunicación. Porcentaje de reactivos con problemas						
Análisis	**Respuesta correcta p<.10 ó p>.9**	**Ptbis<.15**	**Ptbis<.20**	**¿Clave correcta mal especificada?**	**Distractores con Ptbis>0**	
Global	1.7	49.3	65.6	13.8	12.1	
Autónomo	5.1	54.1	65.9	27.3	36.3	
Bachilleratos estatales	5.4	52.7	60.6	39.7	58.3	
CECyTE	1.4	61.4	72.1	22.5	28.5	
Colegio de Bachilleres	3.9	51.0	60.0	25.6	40.0	
CONALEP	2.3	58.6	70.1	24.5	33.0	
DGETA	3.7	57.2	65.4	41.4	59.2	
DGETI	3.9	58.0	72.1	19.2	23.1	
EMSAD	3.9	54.2	63.9	29.0	43.2	
Otros Subsistemas	2.0	42.3	54.1	20.6	29.3	
Privado	3.7	57.2	69.6	34.6	47.3	
Telebach estatales	7.9	46.5	55.2	38.6	62.0	

Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

Tabla 12. Lenguaje y Comunicación. Frecuencia de reactivos con DIF (análisis global)		
Tamaño del DIF	**Hombres**	**Mujeres**
Grande	6	7
Intermedio	13	16
Insignificante	153	160

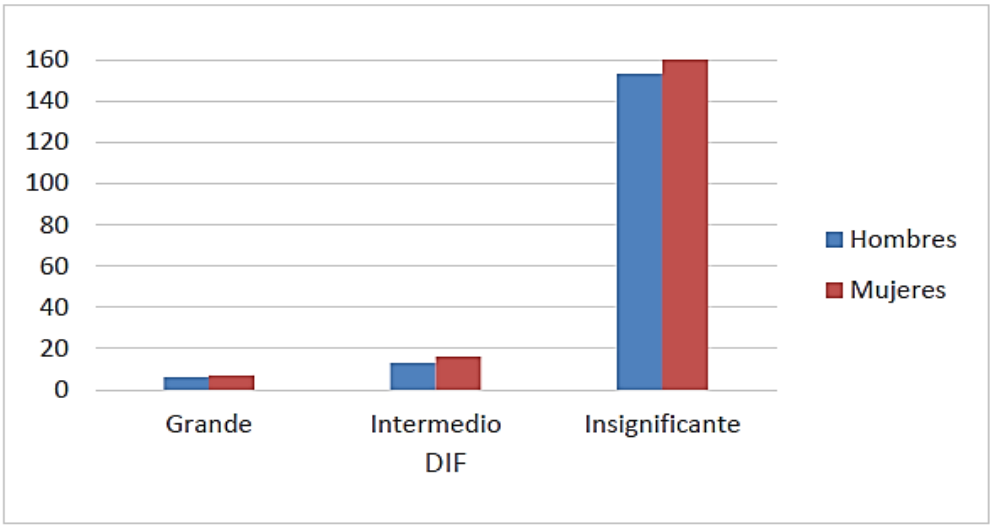
Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

Gráfica 1. Lenguaje y Comunicación. Frecuencia de reactivos con DIF (análisis global)



Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

Gráfica 2. Lenguaje y Comunicación. Frecuencia de reactivos con DIF (análisis global)



Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

MATEMÁTICAS

Tabla 13. Estadísticos clásicos de las formas de Matemáticas (análisis global)

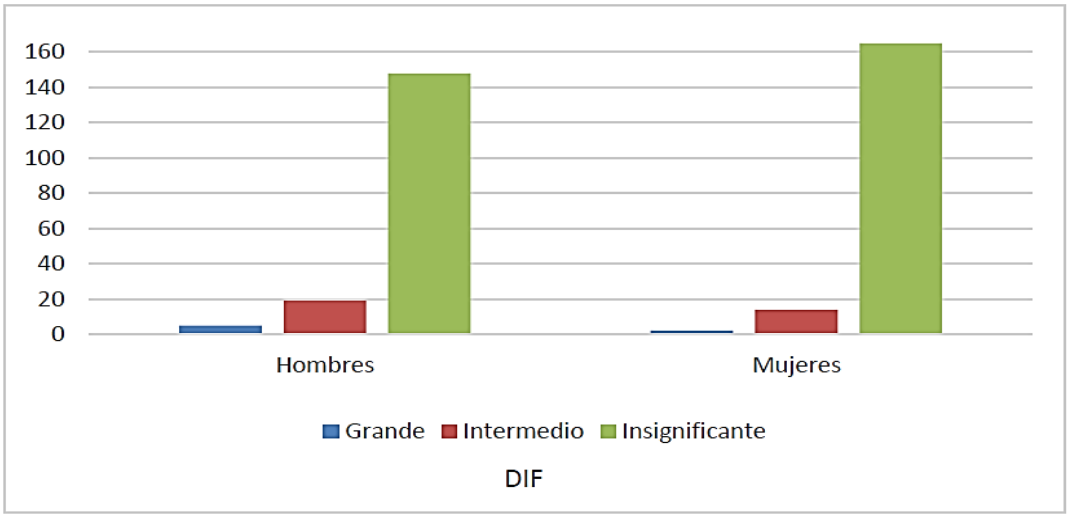
Indicador	Forma 04	Forma 05	Forma 06	Forma 11	Forma 12	Forma 13	Forma 14
N de reactivos	51	51	51	50	50	50	50
N de examinados	764	741	753	501	481	490	516
Mínimo	3	0	7	5	0	4	6
Máximo	43	46	40	38	40	36	42
Media	17.243	18.818	17.969	16.146	14.547	15.822	15.031
Mediana	16	18	18	15	14	15	14
Desviación estándar	5.668	6.383	5.213	5.235	4.4	4.995	4.644
Rango intercuantil	6.75	7.5	7	6	5	6	5
Sesgo	1.206	1.059	0.644	0.919	0.835	0.848	1.435
Curtosis	2.373	1.797	0.548	1.369	2.792	1.173	4.314
KR21	0.657	0.723	0.583	0.613	0.477	0.578	0.523
Coeficiente Alfa	0.698	0.749	0.64	0.6605	0.496	0.595	0.561
SEM	3.113	3.193	3.126	3.05	3.124	3.178	3.079
Media P	33.811	36.898	35.234	32.291	29.094	31.645	30.062
Media correlación punto biserial	16.710	19.902	13.821	15.018	9.478	12.697	11.510

Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*.

Tabla 14. Matemáticas. Porcentaje de reactivos con problemas						
Análisis	**Respuesta correcta p<.10 ó p>.9**	**Ptbis<.15**	**Ptbis<.20**	**¿Clave correcta mal especificada?**	**Distractores con Ptbis>0**	
Global	0.8	52.7	67.7	7.6	7.9	
Autónomo	1.7	50.1	61.2	24.9	40.5	
Bachilleratos estatales	12.7	53.0	60.1	37.4	48.7	
CECyTE	4.8	72.0	80.7	29.2	29.5	
Colegio de Bachilleres	2.3	54.7	65.2	25.2	37.1	
CONALEP	2.8	71.1	82.7	27.8	30.6	
DGETA	7.6	52.7	60.6	41.6	64.3	
DGETI	3.4	52.7	64.9	20.1	28.6	
EMSAD	8.5	64.7	77.1	30.1	39.9	
Otros Subsistemas	1.4	39.4	51.3	16.4	27.2	
Privado	9.1	58.4	68.8	34.8	57.2	
Telebach estatales	5.4	38.2	45.6	34.0	60.3	

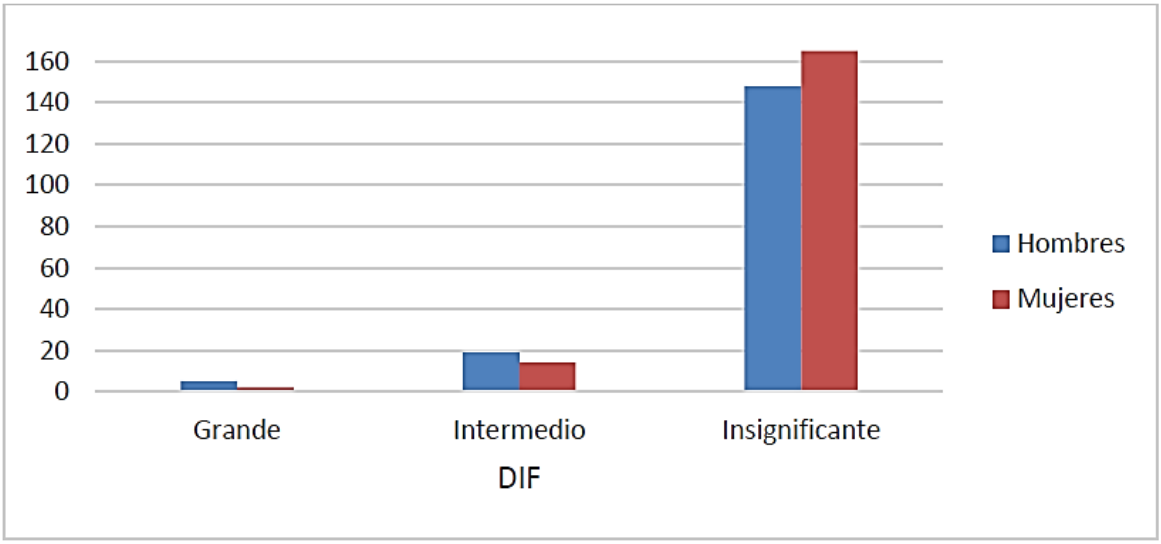
Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12.*

Gráfica 3. Matemáticas. Frecuencia de reactivos con DIF (análisis global)



Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12.*

Gráfica 4. Matemáticas. Frecuencia de reactivos con DIF (análisis global)



Fuente: Dirección General de Medición y Tratamiento de Datos. *Piloteo PLANEA 12*

3.8. Integración de la prueba definitiva

La prueba PLANEA EMS 2017 evaluó el logro del Sistema Educativo Nacional en los grados terminales de la EMS. Este instrumento de evaluación tuvo un diseño matricial, debido a que este formato apoya la evaluación de un mayor número de contenidos curriculares, con lo que se logra una mayor precisión en la medición. Su diseño, entonces, estuvo basado en bloques de reactivos (de 25 ítems cada uno) por cada área disciplinar evaluada. A cada forma de examen, construida a partir de dos bloques, se le denominó "cuadernillo". Este formato de aplicación contuvo 50 reactivos para cada área disciplinar y fue el que se le entregó a los estudiantes. Los ítems se eligieron a partir del universo de reactivos y fueron seleccionados tomando en consideración los criterios estadísticos obtenidos en la prueba piloto, como a continuación se detalla.

3.8.1. Selección de los reactivos de la prueba a partir de los criterios estadísticos

Una vez que los reactivos producidos por el Comité de Elaboración de Reactivos fueron aplicados en la prueba piloto, sus respuestas se integraron en archivos y se analizaron para verificar el comportamiento de cada una de las preguntas en función de su grado de dificultad. También se observó si estas permitían distinguir entre los alumnos que tenían altas probabilidades de responder correctamente y los que no las tenían. Al respecto cabe recordar que se llevaron a cabo dos análisis: el análisis psicométrico de reactivos de logro de opción múltiple y el análisis de comportamiento diferencial de reactivos.

Con estos datos se hizo una primera selección de reactivos. Se eligieron aquellos cuyo Ptbis (Correlación del punto biserial corregida) se localizara en un rango de entre 0.15 y 0.20. Otro criterio de elección fue que los ítems elegidos para integrar la prueba nacional no debían presentar funcionamiento diferencial (DIF) para hombres, mujeres o algún grupo social.

Esto significó que se eligieron, en un primer momento, aquellos reactivos cuyas opciones de respuesta se podían diferenciar claramente (la respuesta correcta y las que no lo eran) y que, por ende, no generaran confusión en los alumnos. Se optó, entonces, por reactivos con los que, a partir de sus resultados, se podía ver si los estudiantes contaban o no con la habilidad requerida para responderlos.

No sólo se eligieron los ítems con mejores valores estadísticos, sino también se seleccionaron reactivos de la mayor cantidad posible de especificaciones para privilegiar la diversidad de contenidos. Se buscó, también, que existiera un número proporcional de reactivos de cada uno de los ejes temáticos de cada área disciplinar.

Finalmente, un criterio no formal, pero que se aplicó, fue el de seleccionar aquellos reactivos con mejores criterios estadísticos para poblaciones vulnerables o tradicionalmente con menores resultados, por ejemplo, los telebachilleratos.

3.8.2. Criterios de equivalencia y distribución de los bloques

Respecto a los criterios de equivalencia para armar los bloques (de 25 reactivos cada uno) cabe señalar los siguientes:

1. Se estableció que existiera un número proporcional de reactivos correspondientes a cada eje temático.
2. Se procuró que el promedio de dificultad de los ítems de cada forma fuera semejante.
3. Se evitó que hubiera reactivos de la misma especificación en el mismo bloque, o por lo menos, se intentó que fueran los menos.
4. Se procuró que los ítems que medían el mismo contenido estuvieran en bloques diferentes.
5. Se proyectó que los reactivos pertenecientes a una misma unidad de evaluación estuvieran juntos y presentados en un nivel progresivo de dificultad.
6. Se cuidó de que en un mismo bloque no hubiera ítems que ayudaran en la solución de otros.

En la prueba de Lenguaje y Comunicación, además:

1. Se hizo una clasificación de los 150 ítems en: Reactivos racimo (anclados a una lectura), Reactivos independientes y Reactivos s.
2. Se buscó dejar un número semejante de lecturas en cada bloque (con un número semejante de Reactivos racimo en cada forma).
3. Se procuró dejar un número semejante de reactivos independientes en cada bloque.
4. Para la forma impresa, el área de diseño adecuó los reactivos atendiendo al espacio en las páginas y acomodó las imágenes para fueran legibles.

En las tablas 15 y 16 se observa la distribución de cada uno de los bloques de reactivos por área disciplinar:

Tabla 15. Distribución de bloques Lenguaje y Comunicación						
Ejes temáticos	Número de reactivos en los bloques					
	A	B	C	D	E	F
Manejo y construcción de la información	9	9	9	9	8	8
Texto argumentativo	5	6	6	5	6	6
Texto expositivo	5	6	5	5	5	5
Texto literario	6	4	5	6	6	6
Total de reactivos	25	25	25	25	25	25
Características de los reactivos	Bloques					
	A	B	C	D	E	F
Promedio grado dificultad de los reactivos	0.506	0.528	0.549	0.511	0.506	0.522
Dificultad de los reactivos que conforman la prueba de Lenguaje y Comunicación EMS 2017						

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 16. Distribución de bloques Matemáticas						
Ejes temáticos	Número de reactivos en los bloques					
	A	B	C	D	E	F
Sentido numérico y pensamiento algebraico	9	9	8	9	9	9
Cambios y relaciones	9	8	9	9	8	8
Forma, espacio y medida	2	3	3	2	2	2
Manejo de la información	5	5	5	5	6	6
Total de reactivos	25	25	25	25	25	25
Características de los Reactivos	Bloques					
	A	B	C	D	E	F
Promedio grado dificultad de los reactivos	0.383	0.384	0.386	0.384	0.388	0.383
Dificultad de los reactivos que conforman la prueba de Matemáticas EMS 2017						

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

3.8.3. Revisión de patrones de respuesta correcta

Acerca de la distribución de las respuestas correctas, se tomó en cuenta los siguientes aspectos:

1. Las respuestas correctas debieron cubrir los incisos A, B, C, D.
2. Tuvo que existir un número equivalente de incisos A, B, C, D, como respuestas correctas.
3. Debió cuidarse que no existiera un patrón lógico de respuestas correctas; es decir, se procuró la aleatoriedad.
4. En función de la edición de la forma impresa, se realizó una última revisión de la distribución de los incisos de las respuestas correctas, para verificar que no existieran más de tres incisos iguales consecutivos.

3.9. Edición, armado e impresión de las pruebas

Una vez que se seleccionaron los reactivos para la prueba definitiva, fue necesario realizar el ensamble respetando los criterios de armado que se señalaron en el apartado anterior. De la misma manera, se cuidó que todos los ítems tuvieran un diseño editorial homogéneo.

A continuación se presenta el procedimiento de edición, armado e impresión de la prueba nacional PLANEA EMS 2017.

3.9.1. Revisión de las impresiones de la prueba

Los criterios establecidos por el INEE para el desarrollo de instrumentos de evaluación señalan que: "Antes de reproducir el cuadernillo [...] se debe revisar su legibilidad, que no tenga errores ortográficos ni de diseño (tamaño y nitidez) y que la posición de los reactivos sea correcta" (INEE, 2017: 29).

3.9.2. Revisión de la versión definitiva de la prueba

Antes de entregar los archivos y las maquetas impresas de los cuadernillos que conformaron el instrumento de evaluación, el personal del INEE realizó una revisión de los cuadernillos que conformaron la versión definitiva de la prueba. Fue importante poner atención nuevamente a las formas del instrumento, para verificar que, efectivamente, contaran con las mismas características técnicas, los reactivos tuvieran estadísticos similares, medidas y desviaciones estándar similares, índices de confiabilidad equivalentes y un número proporcional de reactivos correspondientes a cada eje temático. Además, se verificó que los reactivos siguieran los mismos criterios de diseño editorial y que no tuvieran errores ortotipográficos.

3.9.3. Envío a prensa para generar cuadernillos

Una vez concluido este proceso, se llevó a cabo una reunión con la empresa elegida para imprimir la prueba, cuya selección se realizó de manera reservada para garantizar la confidencialidad de la información. En dicha reunión, el personal del Instituto entregó al proveedor de servicios de impresión el paquete de materiales mediante oficio, la ficha con las especificicaciones y ratificó las fechas para la entrega del pedido.

3.9.4. Revisión de los cuadernillos y autorización para su reproducción

Posteriormente a la reunión con el proveedor de servicios de impresión, y de acuerdo con las fechas establecidas, éste envió una prueba de impresión de los cuadernillos de PLANEA EMS 2017 al INEE, misma que fue revisada al azar para corroborar que, efectivamente, correspondieran con la versión original de la prueba. Esto se llevó a cabo consultando las pruebas de impresión previamente autorizadas y firmadas.

En algunos casos, es posible que se haya hecho modificaciones que fueron detalladas a los servicios de impresión; de ser así, se recibieron nuevamente las pruebas de impresión con las modificaciones y, una vez revisadas nuevamente, éstas fueron aprobadas y firmadas para su reproducción, mediante memorándum.

4. Administración y resguardo del instrumento de evaluación

Los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* (2017), establecidos por el Instituto Nacional para la Evaluación de la Educación (INEE), retoman la conceptualización del American College Testing, que en su documento *Summative Technical Manual* (2016) determina que la administración de un instrumento de evaluación resulta esencial en el proceso, ya que éste es el paso que contribuye a la validez de la interpretación de los resultados, los análisis derivados de ellos y la equidad de la evaluación, siempre y cuando la administración garantice condiciones estandarizadas para la aplicación (INEE, 2017). Aunado a lo anterior, en el estudio *Las pruebas ENLACE y EXCALE. Un estudio de validación* (2015), se observó que nuestro país se enfrenta a un desafío específico: "En México [existe una] complejidad [que] tiene que ver con la forma en que están organizados el sistema educativo nacional y los de las entidades, con las características geográficas y las condiciones sociopolíticas locales" (Martínez, 2015: 68).

Debido a estas particularidades, es necesario detallar cuáles fueron los protocolos de administración y los inventarios para el resguardo de la información que siguió la aplicación definitiva de PLANEA EMS 2017, con la finalidad de dejar constancia de los estándares bajo los cuales se llevó a cabo la aplicación y el resguardo de la información, y para evidenciar que ésta es útil para identificar áreas de oportunidad en la EMS. Asimismo, es necesario ilustrar las irregularidades que se enfrentaron durante esta fase y las decisiones que se tomaron para destacar algunas áreas de mejora con respecto al proceso de aplicación de las pruebas estandarizadas.

Cabe recordar que la Dirección General de Levantamiento y Procesamiento de Datos (DGLPD) del INEE es la encargada de diseñar y realizar los operativos en campo, por lo que la información aquí agrupada proviene del *Reporte técnico. PLANEA EMS, ELSEN 2017*, presentado por esta Dirección. También se consultó el reporte *PLANEA 2017 Educación Media Superior. Diseño muestral* de la Dirección General de Medición y Tratamiento de Datos (DGMTD) que contiene información sobre la definición del marco muestral para la aplicación nacional.

4.1. Aplicación de la prueba nacional

Los días 04 y 05 de abril de 2017 se llevó a cabo la administración de PLANEA EMS 2017, correspondiente a la modalidad referida al Sistema de Educación Obligatoria (PLANEA SEN), de la cual el INEE es responsable. Esta modalidad, como se ha dicho en otras secciones de este Manual:

> Tiene el propósito [general] de aportar información a las autoridades educativas federales, locales, organismos descentralizados y a la sociedad en general, sobre el logro de aprendizajes clave que adquieren los estudiantes del SEN en su conjunto, en su tránsito por la enseñanza obligatoria, respecto a los campos de formación de Lenguaje y Comunicación y Matemáticas (INEE, 2018: 11).

En el caso de PLANEA EMS 2017, su objetivo específico es conocer la medida en que los estudiantes del último grado de educación media superior (EMS) logran el dominio de un conjunto de aprendizajes esenciales en dos áreas disciplinares: Lenguaje y Comunicación y Matemáticas.

Por lo tanto, el levantamiento de datos es un procedimiento que debe procurar la consecución de estos objetivos. Para lograrlo, PLANEA EMS 2017 contó con tres instrumentos de recaudación de información:

Tabla 17. Instrumentos de aplicación de PLANEA EMS 2017

Instrumento	Descripción	Informantes
Prueba de Lenguaje y Comunicación	Busca conocer el desarrollo de las competencias comunicativas a partir del uso y estudio formal del lenguaje.	Estudiantes de grados terminales de educación media superior.
Prueba de Matemáticas	Explora las habilidades para dar solución y argumentación a problemas basados en el razonamiento.	Estudiantes de grados terminales de educación media superior.
Cuestionario para el alumno	Inquiere el entorno personal, familiar y escolar en que se desarrollan los aprendizajes.	Estudiantes de grados terminales de educación media superior.

Fuente: *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD.

Aunado a estos instrumentos:

> la SEP solicitó que los directores respondieran un cuestionario en línea, por lo que en la jornada del día 1 se les entregó la 'Guía del Cuestionario para el Director', URL y clave de acceso a la plataforma para su resolución (DGLPD, 2017c: 7).

Debido a que es una prueba de tipo matricial y muestral, en la que cada alumno respondió alguna de las formas constituidas por 50 reactivos de Lenguaje y comunicación y 50 de Matemáticas, la aplicación de PLANEA EMS 2017 se realizó en dos jornadas escolares: el día 04 de abril de 2017 se aplicaron Lenguaje y Comunicación y el Cuestionario del Alumno, y el día 05 de abril se administró la prueba de Matemáticas, como se detallará más adelante en el protocolo de levantamiento de datos.

4.1.1. Determinación de la muestra matricial

Para llevar a cabo una adecuada organización de la aplicación de PLANEA EMS 2017 fue fundamental seleccionar una muestra representativa de estudiantes de grados terminales de la EMS de los diferentes subsistemas de la Subsecretaría de Educación Media Superior (SEMS) (población objetivo), la cual fue construida con base en la información de dos fuentes principales:

> La primera de ellas fue la Estadística Oficial: Información actualizada al inicio del ciclo 2016-2017, registrada en los centros de trabajo mediante el Formato Estadístico 911, que fue proporcionada a la Subdirección de Muestreo del INEE por la Subsecretaría de Educación Media Superior. La matrícula contenía, además del total de alumnos por modalidad (mixta o escolarizada) según el grado, información sobre los turnos impartidos en cada modalidad, pero no la matrícula por turno.
>
> La segunda fuente fue la información enviada voluntariamente por los centros de trabajo a la Secretaría de Educación Pública (SEP), en la que se reporta directamente el listado de los alumnos matriculados en el grado terminal hasta los primeros días de enero del año en curso. Es pertinente mencionar que para algunos centros de trabajo el listado con los nombres de los alumnos no está disponible, pero sí la cantidad de alumnos del grado terminal (DGMTD, 2017b: 5-6).

Debido a que se trató de un diseñó con cobertura geográfica nacional, cuya finalidad fue obtener información de cada una de las 32 entidades del país, el marco de muestreo, como lo determina el *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD (2017c), fue conformado por 16 034 planteles de EMS de los distintos subsistemas.

> El diseño para la selección de planteles fue polietápico, estratificado y por conglomerados: en la primera y segunda etapas los estratos fueron las entidades federativas y los tipos de sostenimiento, respectivamente. La tercera etapa se realizó por conglomerados dentro de cada centro de trabajo respecto a los turnos, los cuales fueron las Unidades Primarias de Muestreo (UPM).
>
> Así, la muestra obtenida fue de 2 437 planteles-turno, en los que se estimó la participación de 153 289 alumnos en todo el país (DGLPD, 2017c: 5).

A continuación, la tabla 18 muestra el número de planteles que constituyeron la muestra por entidad.

Tabla 18. Distribución de planteles por entidad

Entidad	Número de centros de trabajo-turno
Aguascalientes	63
Baja California	75
Baja California Sur	56
Campeche	57
Coahuila	78
Colima	62
Chiapas	71
Chihuahua	75
Ciudad de México	83
Durango	73
Guanajuato	79
Guerrero	92
Hidalgo	81
Jalisco	124
México	106
Michoacán	75
Morelos	70
Nayarit	71
Nuevo León	92
Oaxaca	75
Puebla	96
Querétaro	72
Quintana Roo	61
San Luis Potosí	52
Sinaloa	105
Sonora	69
Tabasco	77
Tamaulipas	74
Tlaxcala	61
Veracruz	80
Yucatán	65
Zacatecas	67
Total	2 437

Fuente: *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD.

También se muestra los subsistemas por tipo de sostenimiento que fueron considerados para la publicación de resultados. Acerca de lo anterior, la DGMTD expresa que "El día 13 de septiembre de 2017 se solicitó valorar la pertinencia de incorporar y/o fusionar dominios de estudio, así como renombrarlos para la publicación de resultados". (DGMTD, 2017b: 24). El resultado final de este procedimiento es el que se muestra en la tabla 19.

Tabla 19. Subsistemas considerados para la publicación de resultados

Tipo de servicio educativo	Tipo Sostenimiento: público o privado
Bachilleratos de las Universidades Autónomas	Público
Bachillerato estatal	Público
Telebachillerato estatal	Público
Telebachillerato comunitario	Público
Bachilleratos particulares	Privado
Colegio de Bachilleres	Público
CECYTE	Público
DEGETA	Público
DGETI	Público
CONALEP	Público

Fuente: *Reporte Técnico. PLANEA EMS, ELSEN 2017*, de la DGMTD.

De los 2 437 planteles-turno, los que fueron programados para el levantamiento descendieron a 2 419, ya que la muestra inicial se ajustó debido a inconsistencias relacionadas con el marco muestral: falta de grupos en los grados elegidos para la administración del instrumento de evaluación, así como la duplicidad de planteles en el listado de centros de trabajo participantes. Finalmente, de esta medida (2 419 planteles-tuno), los instrumentos se aplicaron en 2 323, lo que arroja una Tasa de Nacional de Aplicación de 96.03% (DGLPD, 2017c). Para más detalles del diseño muestral y su precisión, ver el Anexo VI.

4.1.2. Diseño de la aplicación nacional

El diseño para el levantamiento de datos en una muestra de escuelas y alumnos, como es la de PLANEA EMS 2017, se apegó a los estándares de administración del instrumento, antes, durante y después de la aplicación, para evitar ajustes que pudiesen alterar la medición del constructo.

Un primer paso en la fase de aplicación fue definir la estructura operativa con tareas diferenciadas y ordenadas en función de la consecución de los objetivos propuestos para PLANEA EMS 2017. El *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD (2017c) expresa que:

> Con la finalidad de organizar y distribuir responsabilidades en el levantamiento de datos, la DOC [Dirección de Operación de Campo] diseñó la estructura operativa que fue coordinada por cada una de las Direcciones del INEE en los estados (DINEE); las cuales, a su vez, organizaron al personal participante, por lo que la estructura en cada entidad se determinó en función de la cantidad de planteles de la muestra (DGLPD, 2017c: 8).

Para cumplir con un riguroso control, se contrató personal externo eventual para la administración del instrumento de evaluación. Se agrupó de acuerdo con siete figuras o cargos:

- Apoyo operativo: 35 (1 por cada entidad, salvo Jalisco y México con 2 cada una).
- Coordinador de zona: 212 (de acuerdo con la planeación operativa, en todas las entidades en relación a la cantidad de muestra asignada).

- Aplicador: 4 026 (de acuerdo con la planeación operativa en todas las entidades y en relación a la cantidad de muestra asignada; por capítulo 4000).
- Responsable en digitalización: 32 (1 por cada entidad).
- Operador de digitalización: 32 (1 por cada entidad).
- Operador de almacén: 32 (1 por cada entidad).
- Chofer: 36 (no se contrató a esta figura en Aguascalientes e Hidalgo. En Jalisco, Oaxaca, México, Puebla, Sinaloa y Veracruz se contrató a 2 en cada entidad). El resto de las entidades contrató a 1 chofer (DGLPD, 2017c: 8).

Además de esto, se previó la capacitación del personal responsable de la administración de PLANEA EMS 2017, para informarlo acerca de las características de la prueba y los lineamientos a seguir, a partir de talleres que se desglosarán en el siguiente apartado; todo ello con la finalidad de lograr una estandarización en el proceso de la aplicación.

Otra de las decisiones que se tomaron previamente a la aplicación de PLANEA EMS 2017 fue no enviar el material de los planteles-turno (18 en total) que presentaron inconsistencias. Por ello, las cajas con los materiales de estos planteles-turno:

> fueron abiertas por los apoyos operativos y coordinadores de zona para realizar el conteo de cuadernillos, sin desprender la hoja de respuestas; por lo tanto, el Responsable en digitalización no contó ni digitalizó dichas hojas, las cajas fueron marcadas con la leyenda "escuela no aplicada" y quedaron bajo el resguardo de las DINEE (DGLPD, 2017c: 12).

De acuerdo con lo establecido, el responsable de trasladar los materiales a las escuelas-planteles fue el coordinador quien, el día de la aplicación, hizo el contacto inicial con el responsable del plantel, a quien entregó (previa revisión) los materiales para la aplicación. Una vez concluido esto, se le entregó al aplicador los cuadernillos, los cuestionarios de contexto, las hojas de respuesta y el formato para el control de la aplicación en el aula. Con estos elementos, el aplicador pudo llevar a cabo la preparación de la prueba de Lenguaje y Comunicación y del Cuestionario para el alumno (jornada 1) y de Matemáticas (jornada 2). Para los días en que se administró PLANEA EMS 2017 se generó un protocolo de aplicación, que contempló y abarcó las dos jornadas escolares. Éste se presenta en los apartados subsecuentes.

- **Capacitación de los aplicadores de la prueba nacional**

La capacitación es un requisito para la correcta aplicación del instrumento de evaluación. Por ello, los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, establecidos por el INEE, determinan que: "Es necesario instruir de manera formal al personal responsable de la administración sobre las características del instrumento, los lineamientos a seguir en la administración, así como el código de conducta que deberá respetar" (INEE, 2017: 32), todo con la finalidad de que se reduzca cualquier tipo de sesgo durante la etapa de aplicación.

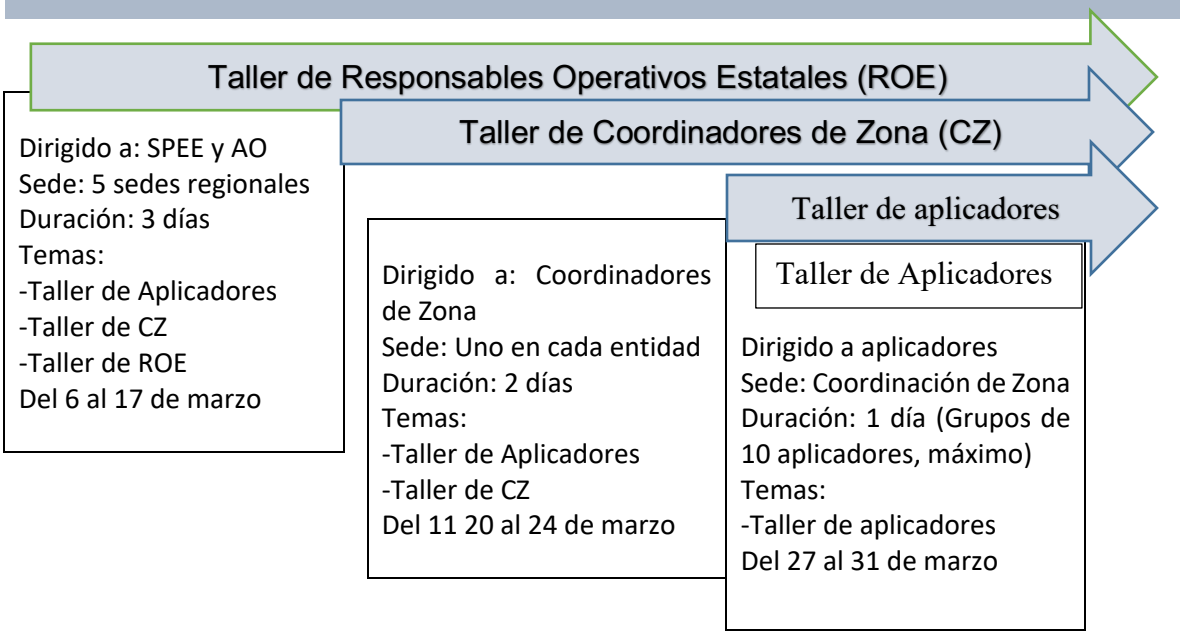
En el caso de PLANEA EMS 2017, la estrategia de capacitación estuvo diseñada en tres talleres: 1. Taller de Responsables Operativos Estatales (ROE), 2. Taller de Coordinadores de Zona (CZ) y 3. Taller de Aplicadores. Cada uno de estos talleres se conformó por actividades e instrumentos que se desarrollaron para dotar al personal operativo de las características técnico-operativas y de las bases del procedimiento para el levantamiento de datos. Por lo tanto, los objetivos de los talleres fueron tres en específico:

- Desarrollar habilidades de instructor en el personal encargado de la réplica de la capacitación en las entidades federativas.
- Lograr la estandarización de la capacitación con la finalidad de reducir posibles sesgos de medición.
- Coadyuvar en la obtención de información con calidad (INEE, 2017: 9).

Además, estos tres talleres contaron con una metodología en cascada, la cual procuró que la información se transmitiera de manera similar y con el menor sesgo posible. Para cumplir este objetivo, la capacitación se dio de manera presencial, de acuerdo al siguiente procedimiento: "personal de la DGLPD capacitó a los apoyos operativos y éstos, a su vez, a los coordinadores de zona, mismos que fueron comisionados para reproducir el Taller de Capacitación de Aplicadores" (INEE, 2017: 9).

En el diagrama 2 se resumen las fases de la capacitación del personal operativo.

Diagrama 2. Etapas de capacitación para aplicadores



Fuente: *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD.

Un segundo proceso de capacitación estuvo orientado a la digitalización. A partir de una herramienta de tipo Ambiente Educativo Virtual (Moodle), se capacitó al Subdirector de Procesos de Evaluación Educativa (SPEE), al Responsable en digitalización, al Operador de digitalización y al Operador de almacén, en sesiones teórico-prácticas se abordaron los procesos de recepción, digitalización, envío y almacenaje de los materiales que constituyeron PLANEA EMS 2017. El Operador de digitalización y el Operador de almacén también fueron aleccionados por el Responsable en digitalización en las DINEE (DGLPD, 2017). Se conformaron cuatro grupos de trabajo, cada uno integrado por ocho entidades de la República mexicana.

La tabla 20 presenta la agenda de capacitación para la digitalización.

Tabla 20. Sedes y fechas de la capacitación para la digitalización		
Grupo	**Entidades participantes**	**Fechas de capacitación**
1	Campeche, Chiapas, Ciudad de México, Guerrero, Oaxaca, Tabasco, Veracruz y Yucatán	27 al 28 de marzo
2	Quintana Roo, Coahuila, Durango, Hidalgo, Nuevo León, San Luis Potosí, Tamaulipas y Zacatecas	28 al 29 de marzo
3	Baja California, Baja California Sur, Chihuahua, Morelos, Puebla, Sinaloa, Sonora, Tlaxcala	29 al 30 de marzo
4	Aguascalientes, Colima, México, Guanajuato, Jalisco, Michoacán, Nayarit y Querétaro	30 al 31 de marzo

Fuente: *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD.

- **Protocolo de levantamiento de datos para la aplicación nacional**

La prueba PLANEA EMS 2017 se aplicó en papel y lápiz, con un cuadernillo con instrucciones, una hoja de respuestas y las respectivas formas constituidas por 50 reactivos de Lenguaje y Comunicación y 50 de Matemáticas. Su administración se llevó a cabo en dos jornadas escolares; en cada una se aplicaron los diferentes instrumentos destinados a recabar la información (pruebas de Lenguaje y Comunicación y Matemáticas, y el Cuestionario para el alumno), tal como se detalla en el siguiente protocolo:

JORNADA 1:

- Organización inicial con el director o con la persona a cargo de la dirección. Duración: 15 minutos.
- Preparación de la aplicación de la prueba de Lenguaje y Comunicación y del Cuestionario para el alumno. Duración: 30 minutos.
- Organización de la aplicación de la prueba de Lenguaje y Comunicación en el aula. Duración: 30 minutos.
- Aplicación de la prueba de Lenguaje y Comunicación. Duración: 150 minutos.
- Organización de la aplicación en el aula del Cuestionario para el alumno. Duración: 15 minutos.
- Aplicación del Cuestionario para el alumno. Duración: 50 minutos.
- Control de datos y material. Duración: 20 minutos.
- Cierre del día uno. Duración: 15 minutos.

JORNADA 2:

- Organización inicial con el director o responsable del plantel. Duración: 10 minutos.
- Preparación de la aplicación de la prueba de Matemáticas. Duración 10 minutos.
- Organización de la aplicación en el aula de la prueba de Matemáticas. Duración: 30 minutos.
- Aplicación de la prueba de Matemáticas. Duración: 120 minutos.
- Control de datos y material de aplicación. Duración: 20 minutos.
- Cierre de la aplicación en el plantel. Duración: 20 minutos (DGLPD, 20017c).

De acuerdo con lo establecido, el contacto inicial el día de la aplicación fue con el responsable del plantel, a quien se le dio el oficio de entrega de materiales y se le solicitó un espacio para su revisión. Posteriormente, al aplicador se le entregaron los cuadernillos de la prueba (Lenguaje y Comunicación o Matemáticas, según correspondiera con el calendario de aplicación), los cuestionarios de contexto del alumno y las hojas de respuesta. El aplicador verificó la que la cantidad

de materiales correspondiera con el número de alumnos y procedió con la preparación de la aplicación de la prueba (de Lenguaje y Comunicación o Matemáticas, según correspondiera), y del Cuestionario para el alumno (sólo en la jornada 1).

Al término de la jornada de aplicación, los materiales fueron contabilizados en presencia del responsable del plantel y fueron sellados y etiquetados para su traslado al INEE, previamente firmados de conformidad por el responsable del plantel.

4.1.3. Protocolo de seguridad para la administración de la prueba nacional

La administración de PLANEA EMS 2017 siguió un estricto protocolo de seguridad, con la finalidad de garantizar la confidencialidad de los materiales y de la información recabada. Esto se debe a la importante incidencia que esta prueba tiene en la toma de decisiones y en la definición de políticas públicas para el mejoramiento de la educación. Por esta razón, las normas que se siguieron durante la aplicación fueron las siguientes:

- Se informó a la comunidad escolar sobre su participación en la administración de PLANEA EMS 2017.
- Sólo el aplicador y los alumnos seleccionados para la muestra pudieron permanecer dentro del aula designada para la administración de PLANEA EMS 2017.
- Con la finalidad de obtener datos puntuales del levantamiento de datos, personal del INEE o personal operativo capacitado para este fin supervisaron la administración de PLANEA EMS 2017.
- El aplicador debió señirse a la guía expresamente diseñada para apoyarlo en su labor.
- Los alumnos que terminaron antes de lo previsto tuvieron que entregar todo su material al aplicador y éste no pudo ser devuelto a los alumnos una vez terminada la resolución de las pruebas.
- Todo el material que no fue utilizado durante la administración de la prueba piloto nacional fue resguardado, en todo momento, por el aplicador.
- Una vez terminada la jornada, el aplicador mantuvo en custodia todos los materiales de la aplicación hasta recibir indicaciones de su coordinador.

Este procedimiento se llevó a cabo durante las dos jornadas de aplicación.

- **Informes de la aplicación nacional**

A continuación se presenta información relacionada con el levantamiento de datos de la prueba nacional, sintetizada en el *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD, y de los resultados de la aplicación en términos de la duración promedio de las actividades de aplicación con respecto al protocolo, el porcentaje de la muestra supervisada por entidad federativa, la tasa nacional de aplicación y la tasa de respuesta por instrumentos.

Con respecto a la duración promedio de las actividades de aplicación con respecto al protocolo, se destaca lo siguiente:

- El promedio de tiempo invertido en la aplicación de la prueba de Lenguaje y Comunicación fue de 55 minutos, menor en relación con el protocolo.

- La aplicación de la prueba de Matemáticas tuvo un promedio de 8 minutos por debajo del establecido en el protocolo de aplicación.
- El promedio de duración de la aplicación del Cuestionario para el alumno también tuvo una diferencia de 8 minutos menos con respecto a lo programado en el protocolo (DGLPD, 2017c) (anexo VI).

Esta información se presenta como referencia en posteriores aplicaciones para los operativos en campo concernientes a PLANEA EMS 2017.

El porcentaje de la muestra supervisada por entidad federativa fue de 11%, ya que:

- De un total de 1 842 centros de trabajo donde se realizó la aplicación de PLANEA EMS [...] 2017, se obtuvo información de 205 [...]. En 200 centros, estructura operativa habilitada realizó el monitoreo y el personal del INEE lo hizo en 5 más (DGLPD, 2017c: 18).

A continuación se presenta la tabla 21, donde se detalla el porcentaje de supervisión en cada entidad federativa participante.

Tabla 21. Distribución porcentual de la muestra de supervisión por entidad federativa

PLANEA EMS, ELSEN 2017 entidad	Porcentaje
Aguascalientes	2.44%
Baja California	3.41%
Baja California Sur	2.93%
Campeche	2.44%
Chiapas	3.41%
Chihuahua	2.44%
Ciudad de México	2.93%
Coahuila	3.41%
Colima	3.41%
Estado de México	4.39%
Guanajuato	2.93%
Guerrero	2.93%
Hidalgo	3.41%
Jalisco	4.39%
Michoacán	3.90%
Morelos	2.93%
Nayarit	2.93%
Nuevo León	4.39%
Oaxaca	3.41%
Puebla	3.90%
Querétaro	3.41%
Quintana Roo	2.44%
San Luis Potosí	1.95%
Sinaloa	3.90%
Sonora	3.41%
Tabasco	2.93%
Tamaulipas	2.44%
Tlaxcala	2.44%
Veracruz	3.41%
Yucatán	3.90%
Zacatecas	3.41%

Fuente: *Reporte técnico. PLANEA EMS, ELSEN 2017*, de la DGLPD.

Este monitoreo permitió obtener resultados que, en futuras aplicaciones, apoyarán la planeación y logística de las operaciones en campo.

Acerca de la tasa nacional de aplicación (anexo VII), el *Reporte técnico* de la DGLPD establece lo siguiente:

- Los planteles-turno en los que se aplicaron los instrumentos ascienden a 2 323, de 2 419 programados para el levantamiento; esto corresponde a una Tasa de Nacional de Aplicación de 96.03%.
- Es importante resaltar que 12 entidades realizaron la aplicación en el total de sus planteles asignados: Aguascalientes, Baja California Sur, Campeche, Chihuahua, Durango, Guanajuato, Guerrero, Querétaro, San Luis Potosí, Tlaxcala, Veracruz y Zacatecas.
- [Mientras que] la Ciudad de México [fue] la entidad con la menor tasa de aplicación, 50.63%; la causa más recurrente de este porcentaje fue que no se permitió la aplicación; esta situación se dio en 15 centros del Instituto de Educación Media Superior de la Ciudad de México (IEMS), en 13 planteles pertenecientes a la Universidad Nacional Autónoma de México (UNAM) (nueve planteles de la Escuela Nacional Preparatoria en ambos turnos y el Colegio de Ciencias y Humanidades), así como en un centro escolar de sostenimiento particular.
- Chiapas [fue] la segunda entidad con el menor porcentaje de aplicación y la única que reportó problemas en este rubro por causa de la CNTE (DGLPD, 2017c: 12-13).

Acerca de la tasa nacional por instrumento, se destaca la siguiente información:

Prueba de Lenguaje y Comunicación (anexo VIII):

> Con un promedio nacional de aplicación de 80.66%, los resultados de esta prueba muestran que San Luis Potosí y Veracruz fueron las entidades con los mejores porcentajes, de 91.80 y 91.18%, respectivamente; estos casos contrastan con los de Nayarit, 67.39%, y la Ciudad de México, 32.65%, que aparecen con los registros más bajos (DGLPD, 2017: 14).

Prueba de Matemáticas (anexo IX):

> Con el mismo patrón de comportamiento que la Prueba de Lenguaje y Comunicación, la aplicación del instrumento de Matemáticas presenta a San Luis Potosí y Veracruz como las entidades con los mayores registros, de 91.55 y 90.96%, respectivamente; por otra parte, Nayarit, 61.42%, y la Ciudad de México, 30.95%, revelan los resultados menos favorables (DGLPD, 2017: 15).

Cuestionario para el alumno (anexo X):

> La tasa de aplicación para este instrumento fue de 80.56%; como se observa en la siguiente gráfica, la distribución de porcentajes se asemeja a las dos pruebas anteriores, ya que San Luis Potosí y Veracruz ocupan los primeros sitios, con 92.65 y 91.58% respectivamente; situaciones contrarias fueron las de Nayarit, 66.22%, y Ciudad de México, 32.97%, que se ubican en los últimos lugares (DGLPD, 2017: 16).

4.1.4. Protocolo para el resguardo de la información

De la misma manera que en la prueba piloto se resguardó la información siguiendo pasos establecidos, así en la aplicación nacional de la prueba PLANEA EMS 2017 se obedecieron las siguientes pautas para garantizar la seguridad y confidencialidad de la información obtenida en el levantamiento de datos:

- El responsable operativo entregó los instrumentos de evaluación en cajas selladas (una caja por área disciplinar) a cada coordinador, quien verificó que no hubiera etiquetas rotas y, de haberlas, lo reportó inmediatamente.
- El día de la aplicación, el coordinador entregó los materiales de PLANEA EMS 2017 en las correspondientes escuelas-plantel. El director de cada escuela fue quien recibió los materiales en un primer momento, los revisó y verificó nuevamente que los sellos se encontraran en perfectas condiciones.
- Los materiales le fueron entregados al aplicador, quien corroboró que correspondieran, puntualmente, al número de alumnos en el grupo a su cargo.
- Al final de la jornada, los materiales se contabilizaron en presencia del director y el coordinador empaquetó, etiquetó y selló las hojas de respuestas (utilizadas y no utilizadas), anexando las firmas de los responsables.
- El coordinador trasladó las cajas y las entregó al responsable operativo.

Sobre este rubro el *Reporte técnico* de la DGLPD expresa que, una vez reunido el material de la aplicación, se procedió a su integración, a partir de las siguientes acciones: "el Apoyo operativo registró la información de cada formato en el Sistema de Planeación, Seguimiento y Control Operativo (SIPSCO); y el Responsable en digitalización desarrolló sus actividades de acuerdo con las especificaciones emitidas por la Subdirección de Procesamiento" (DGLPD, 2017: 17).

De la misma manera que la prueba piloto, PLANEA EMS 2017 complementó estos procedimientos con un inventario donde se asentó la cantidad de material que se tenía bajo resguardo, con la finalidad de llevar un estricto control de éste y facilitar su localización.

5. Análisis de resultados de instrumentos de evaluación

Los datos por sí solos tienen una limitada importancia; sólo adquieren significado en la medida en que son analizados e interpretados para dar respuesta a las interrogantes planteadas. De ahí se deriva la importancia de exponer el análisis e interpretación que se realizó con los resultados arrojados por PLANEA EMS 2017, ya que, a partir de su relación con otros conceptos, adquirieron un significado pertinente con los objetivos de la prueba.

En esta fase, por lo tanto, se mencionan los elementos considerados para el análisis estadístico del instrumento y los reactivos que lo conforman, así como el modelo de puntuación empleado. Los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* especifican la importancia de conocer las propiedades métricas de los reactivos, ya que esto: "permite identificar los que son aptos para ser considerados en la puntuación de los sujetos, así como aquellos que no aportan información relevante o son defectuosos, por lo que deben ser excluidos" (INEE, 2017: 36), todo esto con el objetivo de aportar evidencia para la validez de las inferencias que se desean obtener a partir de los datos recopilados.

El análisis estadístico fue responsabilidad de la Dirección General de Medición y Tratamiento de Datos (DGMTD) pues tiene entre sus funciones "realizar el análisis y el tratamiento de los datos correspondiente a las mediciones y evaluaciones" (DOF, 31 enero 2017: 108). Los datos que se presentan en este apartado corresponden a los informes técnicos de esta Dirección del Instituto Nacional para la Evaluación de la Educación (INEE).

5.1. Calibración

Antes de la fase de calibración, la DGMTD llevó a cabo una depuración previa de las claves de respuesta, mediante un análisis de los reactivos, el cual tuvo por objetivo verificar el comportamiento de cada una de las preguntas que se incluyeron en los análisis definitivos. Este procedimiento se realizó usando dos modelos: la Teoría de respuesta al ítem (IRT, por sus siglas en inglés) y la Detección de funcionamiento diferencial de reactivos (DIF). Mediante estos análisis se determinó clasificarlos de acuerdo con los problemas que presentaron: discriminación, ajuste, forma de la curva característica y funcionamiento diferencial. La DGMTD tomó la decisión de eliminar dos reactivos por cada área disciplinar (Lenguaje y Comunicación y Matemáticas).

Enseguida, la DGMTD continúo con la calibración de reactivos, análisis que presentó en el documento denominado *Metodología de escalamiento de PLANEA 2017 Educación Media Superior para la Evaluación de Logro en el Sistema Educativo Nacional (SEN),* del cual se presenta el extracto correspondiente:

> La fase de calibración de los instrumentos consistió en la estimación de los parámetros de dificultad de los reactivos para el modelo logístico simple de Rasch.
>
> Para la calibración, el modelo condicional de respuesta al ítem es usado en conjunción del modelo de la población, pero sin utilizar variables de condicionamiento. Más adelante se habla sobre dichas variables.

El método utilizado para hacer la estimación de los parámetros de los reactivos fue el de máxima verosimilitud marginal (Bock & Aitkin, 1981), en el cual se asume independencia condicional entre las respuestas a los reactivos y que la habilidad es un efecto aleatorio. De esta forma, es posible eliminar los parámetros de las personas del proceso de estimación integrando sobre la habilidad, lo cual se reduce a maximizar el logaritmo de la verosimilitud que surge de

$$p(\boldsymbol{X}|\boldsymbol{\alpha},\boldsymbol{\delta}) = \prod_{i=1}^{N} \int p(\boldsymbol{x}_i|\theta,\boldsymbol{\delta}) p(\theta|\boldsymbol{\alpha})\, d\theta$$

La calibración de cada una de las escalas se realizó independientemente de los parámetros de las otras, se hizo utilizando la máxima cantidad información disponible, es decir, con toda la muestra de los estudiantes (DGMTD, 2018).

5.2. Limpieza y validación de datos

La limpieza y validación de datos constituyó un paso previo a los análisis estadísticos, el cual se centró en detectar las inconsistencias que pudieran existir. La DGMTD describe el siguiente procedimiento para llevar a cabo esta tarea:

> El proceso de conformar bases de datos digitales del ámbito educativo a partir de cuadernillos aplicados en centros escolares es susceptible a datos incorrectos, valores perdidos e inconsistencias en los mismos debido a múltiples fuentes de error (impresión, distribución, recolección, aplicación, digitalización, etc.). Con la intención de mejorar la calidad de la información se realiza el proceso de limpieza y validación de datos, previamente a cualquier análisis estadístico, para:
>
> - Detectar errores
> - Detectar inconsistencias
> - Corregir los datos cuando es posible o asignar códigos de identificación
> - Asignar el tipo y ancho más adecuado a las variables
> - Integrar todo en una base de datos consistente
>
> Al final del proceso se cuenta con datos confiables que permitan un análisis eficiente y sin problemas. De no realizarse el proceso de validación y limpieza de datos se corre el riesgo de obtener resultados inconsistentes en los análisis posteriores.

Tabla 22. Códigos de errores

Código	Descripción	Tipo de evaluación
94	No se presentó a la evaluación a la que pertenece este reactivo.	Logro y contexto
97	Reactivo no administrado (no se incluyó en su cuadernillo de examen).	Logro
98	Respuesta múltiple.	Logro y contexto
99	Respuesta omitida.	Logro y contexto

Fuente: Dirección General de Medición y Tratamiento de Datos.

5.3. Análisis estadístico

Para determinar, con base en el modelo psicométrico seleccionado y los estándares establecidos en el protocolo, si los ítems de la prueba PLANEA EMS 2017 se acotaron a los criterios estadísticos de calidad, se propusieron resultados del análisis cuantitativo para determinar qué ítems eran métricamente adecuados; esto se llevó a cabo con la participación de la DGMTD, Dirección que aportó los valores de confiabilidad y el error de medida a través de los análisis que se presentan a continuación, y que fueron tomados directamente de *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, con fecha de agosto de 2017, y *Metodología de escalamiento de PLANEA 2017 Educación Media Superior para la Evaluación de Logro en el Sistema Educativo Nacional (SEN)*.

5.3.1. Modelo de puntuación de las respuestas

El análisis de los reactivos de PLANEA EMS 2017 se reportó a través de dos indicadores:

- Correlación punto biserial (r_{pb}).
- Ajuste al modelo de Rasch (infit).

El primer indicador, la correlación punto biserial (r_{pb}), se trata de:

> [...] una correlación entre una variable dicotómica y una continua. El valor numérico que tenga la correlación dependerá del grado de asociación lineal entre ambas variables. Estrictamente hablando, la correlación punto biserial es un caso particular de la correlación de Pearson en la cual se supone que ambas variables son continuas. En el caso específico del análisis de pruebas, la variable dicotómica está representada por la respuesta de un reactivo la cual puede tomar dos valores: acierto (1) o fallo (0); y la variable continua corresponde al puntaje total obtenido por los examinados.
> Idealmente, la correlación de Pearson puede tomar valores de -1 a 1 en donde estos valores extremos denotan correlaciones perfectas. El valor 1 se obtiene cuando ambas variables mantienen constante la misma tasa de crecimiento lo cual estaría representado por una línea recta en una gráfica de dispersión. Por su parte el valor -1 se obtiene cuando la tasa de crecimiento es constante pero negativa, lo que indica que cuando una variable se incrementa la otra decrece. La correlación punto biserial, por ser un caso especial de la Pearson difícilmente tendrá estos valores extremos aún en el caso ideal pues la tasa de crecimiento o decrecimiento no puede permanecer constante cuando una de las variables es dicotómica.
> Por este motivo, su interpretación se ha cargado al lado de lo que los especialistas en estadística llaman reglas de dedo (*rule thumb*) que surgen del trabajo empírico y que dan respuesta a problemas específicos donde es necesario un juicio. Comúnmente se considera que una r_{pb} en $[0.30, 0.60]$ es adecuada para una prueba, sin embargo también hay que considerar que este rango solamente es válido para preguntas de una dificultad promedio (DGMTD, 2017c: 4).

Sin embargo, hay que considerar que este rango solamente es válido para preguntas de una dificultad promedio: la r_{pb} tiende a decrecer cuando las preguntas son muy fáciles y especialmente cuando son difíciles. Entre los rangos que comúnmente se utilizan para la inclusión de reactivos de opción múltiple, la ACER (*Australian Council in Educational Reasearch*), utiliza una escala r_{pb} de al menos 0.20. Otro indicador es el establecido por Jerard Kehoe en su estudio titulado *Basic item*

analysis for multiple-choice tests. Practical Assessment, Research & Evaluation de 1995, en el que indica que si un reactivo tiene un r_{pb} por debajo de 0.15, éste posee un mal comportamiento, probablemente porque no es confuso o no mide la misma habilidad que los demás ítems. Pero independientemente de estos valores, la DGMTD del INEE establece que:

> [...] debe considerarse el hecho de que cuando se estima cualquier estadístico, los valores tienen una distribución propia que podría considerarse como 'normal' o esperada. Lo importante en el análisis de pruebas, es que solamente unos cuantos reactivos se salgan de estas reglas.
> Dadas estas circunstancias, el INEE propuso incorporar a los análisis de las pruebas solamente aquellos reactivos con una r_{pb} mayor o igual a 0.15 pues si tienen este estadístico deficiente seguramente tendrá comportamientos anómalos en otros estadísticos especializados en el análisis de pruebas (DGMTD, 2017c: 4).

El segundo indicador, el ajuste al modelo de Rasch (*infit*) es un estadístico que:

> [...] indica qué tan bien se aproxima el comportamiento empírico del reactivo al modelo de Rasch. A diferencia de otros modelos de la Teoría de respuesta al ítem (IRT, en inglés), el modelo de Rasch define un modelo de medición en el cual el objeto a medir y el objeto con que se mide son independientes, así que, si se desea utilizar éste modelo, los reactivos tienen que ajustarse a él y no el modelo al comportamiento del reactivo como sucede con los demás modelos de la TRI.
> Idealmente, el infit visto como una variable aleatoria tiene una esperanza de 1 y algunos especialistas consideran que los reactivos tienen un buen ajuste cuando su infit está en el intervalo (0.8,1.2).
> Sin embargo, Linacre relaja dicho criterio al considerar que desajustes por arriba de 2.0 ya tienen efectos negativos en el análisis de las pruebas y por consiguiente se infiere que por debajo de ese valor no afectan de forma importante a los resultados. Wright por su parte, contempla la posibilidad de salirse un poco del intervalo (0.8,1.2) en función de los objetivos y naturaleza de la prueba. Así, al parecer y a juicio de los principales exponentes del modelo de Rasch, este estadístico puede considerarse con relativa holgura (DGMTD, 2017c: 5).

Finalmente, una vez que se determinó qué ítems se incluirían en PLANEA EMS 2017 y cuáles sustentantes cumplieron con los criterios ineludibles para llevar a cabo la estimación, el modelo de Rasch se ajustó: "bajo el supuesto de que la habilidad o nivel de logro sigue una distribución normal multivariada, es decir, se ajustó un modelo bidimensional, con ambas asignaturas. La finalidad fue determinar la factibilidad de usar un modelo bidimensional" (DGMTD, 2017c: 5).

Cabe resaltar que todos los indicadores se obtuvieron utilizando el programa ConQuest V2.

5.3.2. Reporte preliminar del análisis

Antes del análisis definitivo, fue necesario un reporte preliminar de PLANEA EMS 2017 elaborado por la DGMTD, con la finalidad de verificar el comportamiento de cada una de las preguntas que se incluyeron en los análisis definitivos.

En ese "análisis preliminar" se rectificaron algunas claves de respuesta y se eliminaron algunos reactivos, tanto de la prueba de Lenguaje y Comunicación como de Matemáticas. El informe se dividió en dos partes:

- Análisis con un modelo de Teoría de respuesta al ítem.
- Detección de funcionamiento diferencial de reactivos.

A continuación se presenta el reporte preliminar que presentó la DGMTD ante el INEE:

> El modelo aplicado en PLANEA SEN es una forma generalizada del modelo logístico simple de Rasch, este modelo se conoce como modelo logit multinomial con coeficientes aleatorios, que se describe en (Adams, R. J., Wilson, M. and Wang, W.C., 1997). Se implementa mediante el software ConQuest 2 (Adams & Wu, 2002). En este modelo los reactivos se pueden describir por un conjunto de parámetros fijos, desconocidos y se asume que la habilidad de los sustentantes tiene una distribución de probabilidad [...].
> El modelo logístico simple predice la probabilidad de responder correctamente un reactivo, basado en la habilidad del sustentante i y la dificultad del reactivo δ_j. Es decir,
>
> $$p(x_{ij}|\theta_i, \delta_j) = \frac{\exp\left[x_{ij}(\theta_i - \delta_j)\right]}{1 + \exp\left(\theta_i - \delta_j\right)} \ldots (2)$$
>
> Donde:
>
> x_{ij} Es la respuesta del estudiante i al reactivo j, 1 si es correcta y 0 si es incorrecta;
> θ_i Es el nivel de logro del estudiante i en la escala *logit*;
> δ_j Es el parámetro de localización del reactivo j, caracterizando la dificultad del reactivo.
> La indeterminación de los parámetros del modelo se resuelve mediante la imposición de la siguiente restricción.
>
> $$\sum_{j=1}^{n} \delta_j = 0$$
>
> En donde n es el total de reactivos de la prueba (DGMTD, 2018).

Acerca del reporte del análisis realizado con el modelo logístico de Rasch, la DGMTD presentó el siguiente informe:

> En el grado terminal de bachillerato, para cada una de las asignaturas (Lenguaje y Comunicación y Matemáticas) se ajustó el modelo logístico de Rasch de manera totalmente independiente. Una vez hecho esto, el análisis consistió en revisar principalmente tres indicadores:
>
> - Correlación punto biserial (r_{pb})
> - Ajuste al modelo de Rasch ($infit$)
> - Curva característica del reactivo
>
> Los reactivos fueron clasificados con base en estos indicadores de la siguiente forma (DGMTD, 2018: 6-7).

Clasificación de los reactivos según su poder de discriminación de acuerdo a la correlación punto biserial de la respuesta correcta

Rango de discriminación	Clasificación
$r_{pb} \geq 0.15$	El reactivo discrimina de forma aceptable
$r_{pb} < 0.15$	El reactivo debe ser eliminado de la prueba

Clasificación de los reactivos según el ajuste del modelo logístico de Rasch

Rango de discriminación	Clasificación
$0.8 \leq infit \leq 1.2$	El reactivo ajusta de forma aceptable
$infit < 0.8$ o $1.2 < infit < 2$	El reactivo desajusta moderadamente
$infit \geq 2$	El reactivo debe ser eliminado de la prueba

Clasificación de los reactivos en cuanto a la curva característica

Rango de discriminación	Clasificación
La curva característica formada con los datos observados está cerca de la curva característica esperada	El reactivo ajusta de forma aceptable
La curva característica formada con los datos observados no se parece a la curva característica esperada	El reactivo debe ser eliminado de la prueba

Sobre el análisis de detección de funcionamiento diferencial de los reactivos mediante el modelo logístico de Rasch, la DGMTD reportó lo siguiente:

> Para la prueba PLANEA SEN 2017 de nivel bachillerato, el análisis DIF que se realizó fue para los grupos de estudio definidos por la variable sexo, donde el grupo focal o minoritario fue la subpoblación de mujeres y el grupo de referencia fue la subpoblación de hombres; tanto para la asignatura de Lenguaje y Comunicación como para Matemáticas.
>
> Es importante señalar que el tamaño de las subpoblaciones se relaciona directamente con la potencia del análisis DIF de la prueba. Por lo que, previo a éste, se verificó que el tamaño de las subpoblaciones, tanto focal como de referencia, sea al menos de tamaño 500 sustentantes (Alice Guilleux et al., 2014).
>
> El método de Rasch para detectar DIF consistió en la estimación de los parámetros de dificultad de manera aislada en cada uno de los grupos de interés. La diferencia media en logro entre ambos grupos se controló fijando la dificultad promedio de los reactivos en cero:

$$\sum_{j=1}^{n} \hat{\delta}_{j,R} = \sum_{j=1}^{n} \hat{\delta}_{j,F} = 0$$

Se calculó la diferencia en la escala *logit* entre las dificultades de reactivos estimadas para ambos grupos, como una medida para determinar el Funcionamiento Diferencial de Reactivos:

$$\hat{\Delta}_j = \hat{\delta}_{j,R} - \hat{\delta}_{j,F}$$

Se calculó el índice estandarizado que está dado por:

$$Z_j = \frac{\hat{\Delta}_j}{\sqrt{(SE_{\hat{\delta}_{j,R}})^2 + (SE_{\hat{\delta}_{j,F}})^2}}$$

Bajo la hipótesis de que no se presenta DIF, el valor esperado de $\hat{\Delta}_j$ es 0, y Z_j tiene media 0 y varianza 1. Por lo cual, se reportan solamente los reactivos que presentan una diferencia significativa, es decir, aquellos en donde: $Z_j \leq -1.96$ o $Z_j \geq 1.96$.

Una vez identificados los reactivos con Funcionamiento Diferencial significativo, se procede a clasificarlos en tres categorías, bajo el siguiente criterio (Wilson, 2005: 167):

Categoría	Criterio
DIF despreciable "A"	$\lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert < 0.426$
DIF intermedio "B"	$0.426 \leq \lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert < 0.638$
DIF grande "C"	$\lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert \geq 0.638$

En los casos que la diferencia es significativa, es decir, se presenta DIF, el grupo focal de cada análisis se ve favorecido si $\hat{\Delta}_j$ y Z_j son valores positivos (Wilson, 2005: 7-8).

Una vez terminados estos procedimientos, se determinaron los reactivos a eliminar, principalmente aquellos que presentaron problemas, ya sea de discriminación, ajuste, forma de la curva característica o funcionamiento diferencial.

Asignatura	Cantidad de reactivos eliminados	Cantidad de reactivos utilizados
Lenguaje y Comunicación	2	148
Matemáticas	2	148

Fuente: DGMTD, *Metodología de escalamiento de PLANEA 2017.*

- **Análisis de los puntajes obtenidos en los reactivos**

Uno de los análisis que se realizó fue un ajuste del modelo de Rasch, que consistió en un modelo bidimensional con los dos campos disciplinares que constituyeron PLANEA EMS 2017. Esto se realizó: "bajo el supuesto de que la habilidad o nivel de logro sigue una distribución normal multivariada [...]. La finalidad fue determinar la factibilidad de usar un modelo bidimensional" (PLANEA, 2017: 5). En la tabla 23 se detallan los índices de confiabilidad del modelo unidimensional en comparación con el modelo bidimensional.

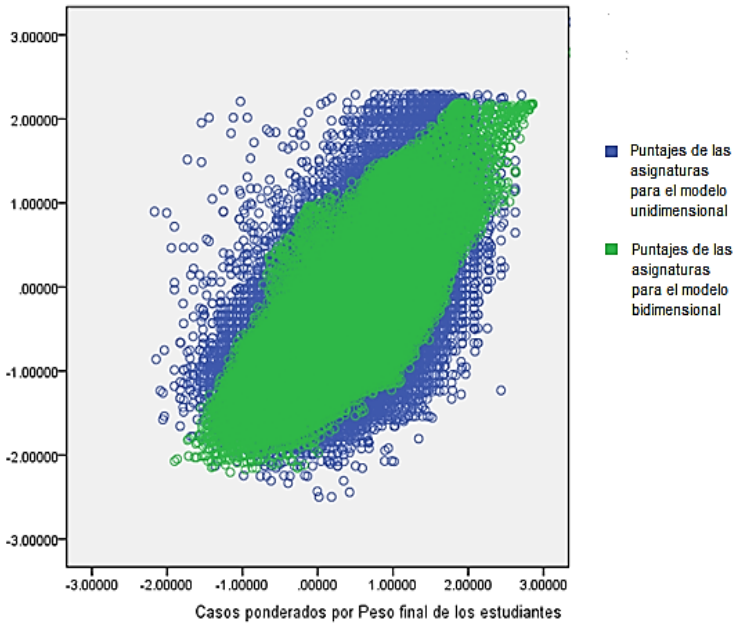
Tabla 23. Confiabilidad de los modelos unidimensional y bidimensional		
Asignatura	Confiabilidad de la prueba en el modelo unidimensional	Confiabilidad de la prueba en el modelo bidimensional
Lenguaje y Comunicación	0.817	0.840
Matemáticas	0.897	0.898

Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Como se puede observar en la relación de la tabla, la confiabilidad incrementó tanto para la asignatura de Lenguaje y Comunicación como para la de Matemáticas.

"Con la finalidad de verificar que existe una alta correlación entre los puntajes de ambas asignaturas, se obtuvieron estimaciones de las habilidades con los modelos unidimensionales o marginales, y posteriormente con el modelo bidimensional" (PLANEA, 2017: 6), como muestra la gráfica 5.

Gráfica 5. Puntajes obtenidos entre asignaturas



Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Como se observa en la gráfica, los puntajes con el modelo bidimensional (en color verde) están menos dispersos que los producidos con los modelos unidimensionales (en color azul). Para verificarlo se calculó la correlación entre los puntajes de Lenguaje y Comunicación y Matemáticas, que fue de 0.725. A partir de ello, se determinó que se utilizaría un modelo bidimensional para PLANEA EMS 2017, porque dicho modelo:

resultó útil para calcular las habilidades con la metodología de los valores plausibles, puesto que permitió estimar la habilidad de los sustentantes en ambas asignaturas aun con una tasa baja de respuesta en los reactivos que sirvieron para estimar el logro (PLANEA, 2017: 6).

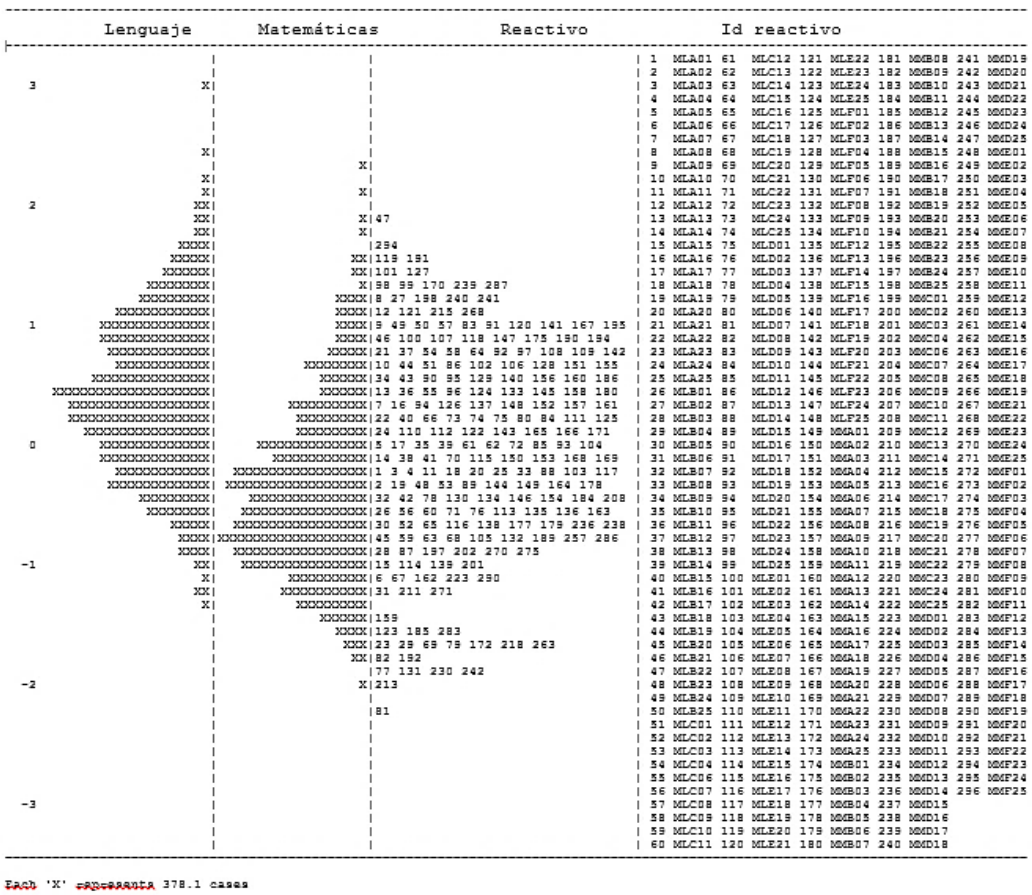
Para mayores detalles del escalamiento de logro del PLANEA EMS 2017, ver el anexo XII.

Correlación	Matemáticas
Lenguaje y Comunicación	0.725

5.4. Reporte final del análisis

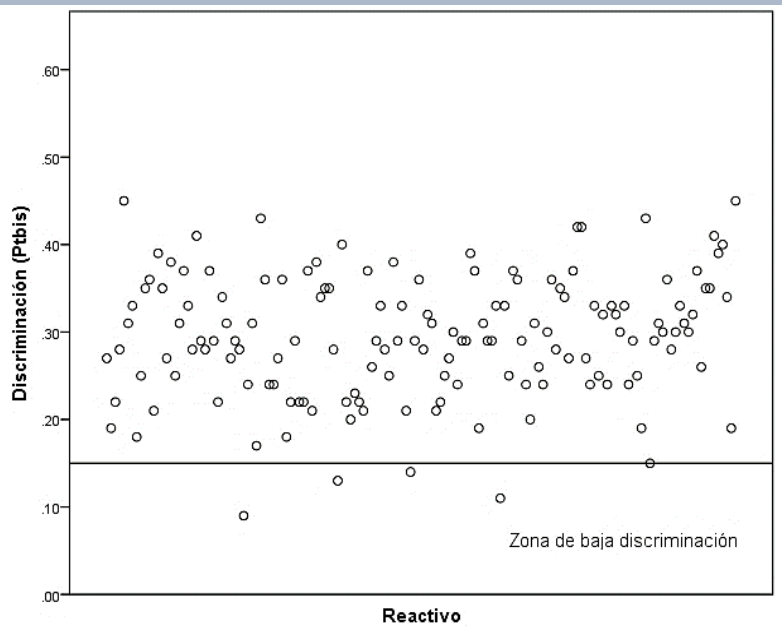
Como se mencionó anteriormente, el análisis de reactivos se elaboró utilizando un modelo de Rasch bidimensional. En las gráficas 6 a 12 se presentan los gráficos y resultados de dicho análisis.

Gráfica 6. Mapa de Wright bidimensional de las pruebas de Lenguaje y Comunicación y Matemáticas



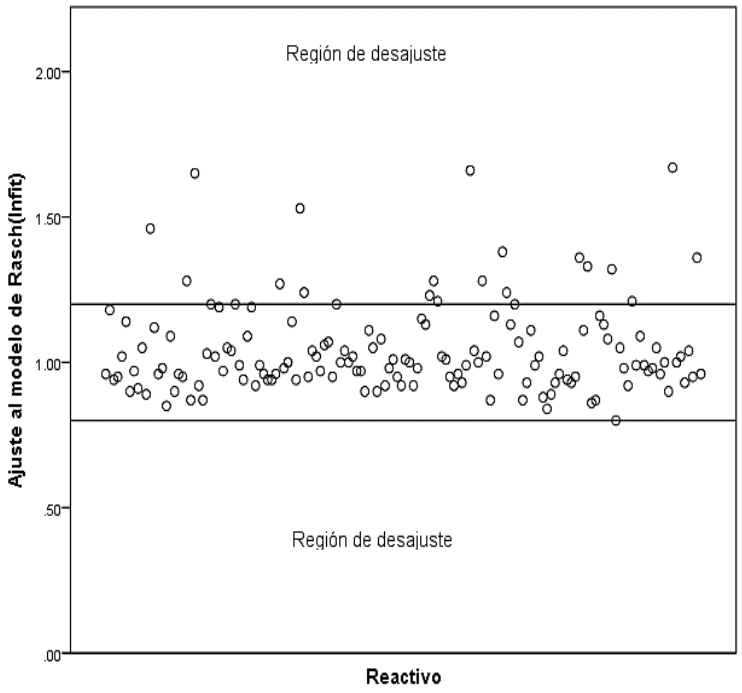
Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 7. Distribución de la correlación r_{pb} de los reactivos de Lenguaje y Comunicación



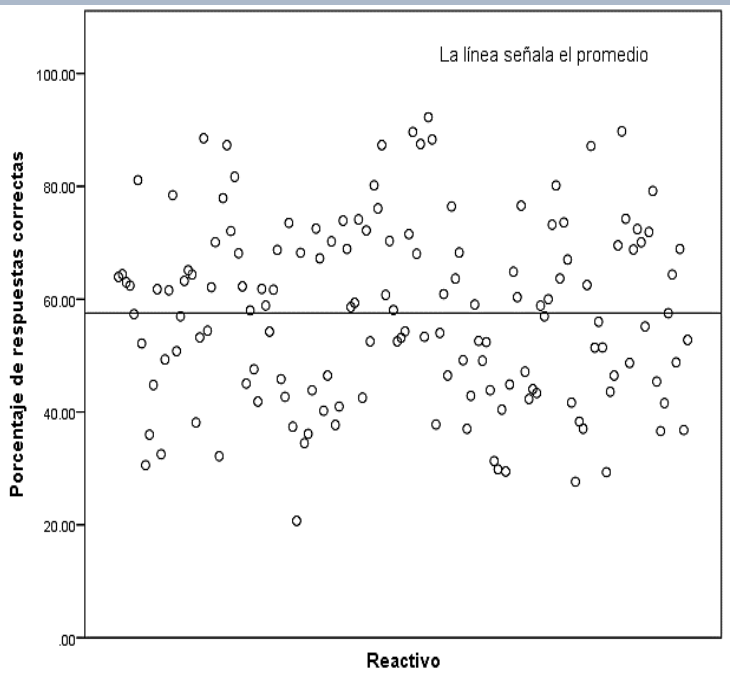
Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 8. Distribución del ajuste al modelo de Rasch (*Infit*) de los reactivos de Lenguaje y Comunicación



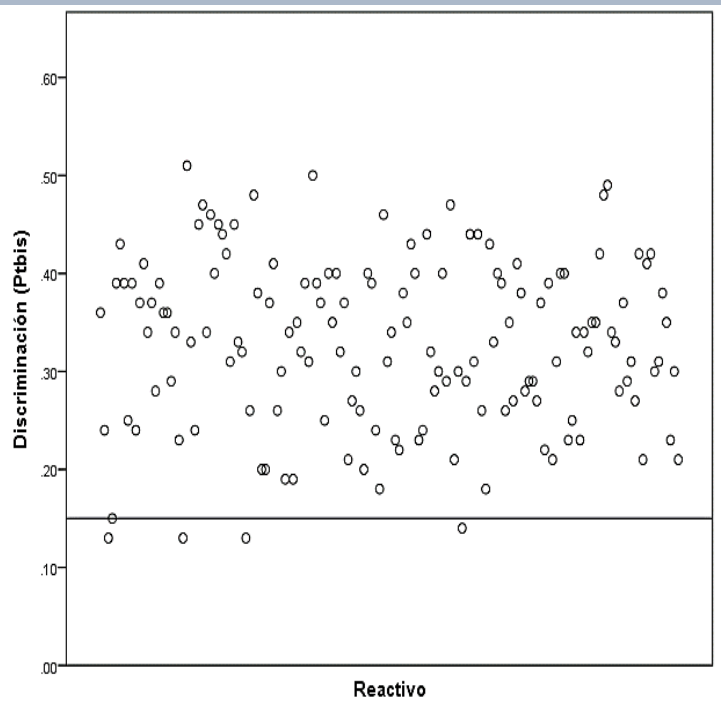
Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 9. Distribución de la proporción de respuestas correctas de los reactivos de Lenguaje y Comunicación



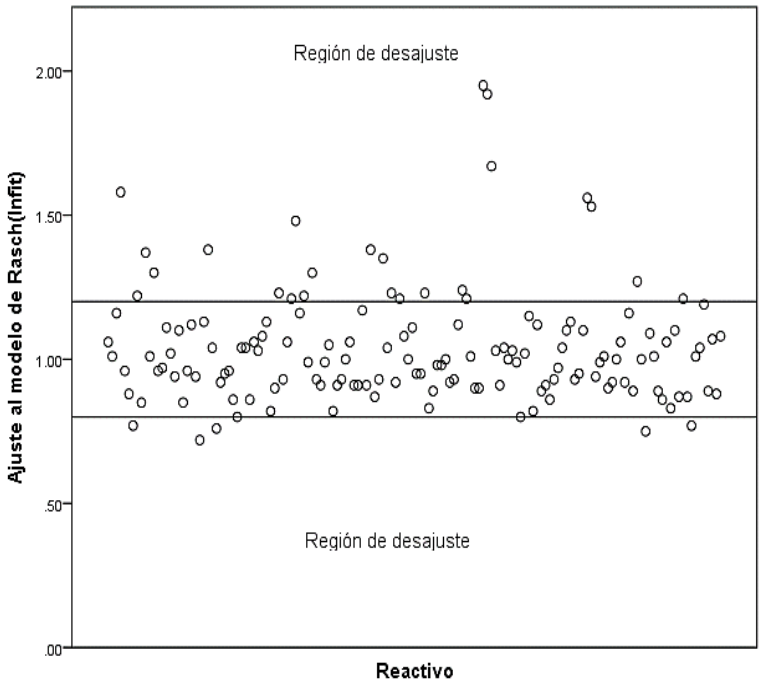
Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 10. Distribución de la correlación r_{pb} de los reactivos de Matemáticas



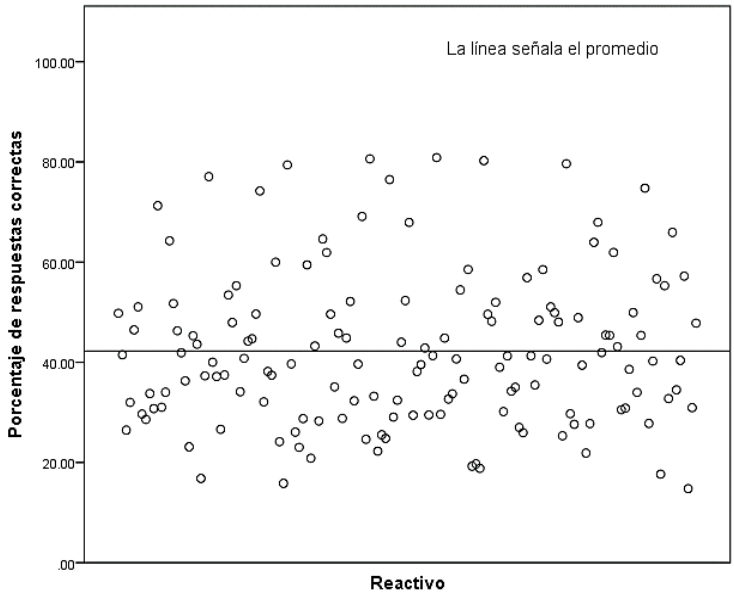
Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 11. Distribución del ajuste al modelo de Rasch (Infit) de los reactivos de Matemáticas



Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Gráfica 12. Distribución de la proporción de respuestas correctas de los reactivos de Matemáticas



Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

Tabla 24. Confiabilidad de las pruebas		
Estadísticos	**Lenguaje y Comunicación**	**Matemáticas**
Cantidad de reactivos	148	148
Cantidad alumnos muestreados (n)	107289	107289
Confiabilidad (EAP/PV RELIABILITY) del modelo bidimensional	0.840	0.898

Fuente: *Análisis de Reactivos PLANEA 2017 Educación Media Superior*, de la DGMTD.

5.5. Modelo de equiparación de resultados en el tiempo

De acuerdo con el INEE, la equiparación es un proceso estadístico que se utiliza para ajustar las puntuaciones de las formas de un mismo instrumento, lo cual permite que las puntuaciones de una forma a otra sean utilizadas de manera intercambiable (INEE, 2017). Esto se lleva a cabo con la finalidad de que exista una línea histórica de los resultados de PLANEA. Para esto se determina una cantidad de "reactivos ancla", los cuales serán comunes a diversas formas del PLANEA EMS, con el propósito de llevar a cabo la equiparación.

6. Difusión y uso de los resultados del instrumento de evaluación

Acerca de la difusión de los resultados de las evaluaciones de logro educativo, el Instituto Nacional para la Evaluación de la Educación (INEE) ha expresado que: "La tarea de evaluación no consiste sólo en producir información relevante, sino también en difundirla amplia y oportunamente entre los actores interesados ofreciendo orientaciones precisas para que los resultados se comprendan, interpreten y utilicen de manera adecuada" (INEE, 2016b: 35). Por ello, la difusión de los resultados también adquiere relevancia dentro del cuerpo expositivo de este Manual técnico, por lo que se precisa describir la interpretación y difusión de los resultados arrojados por la aplicación nacional de PLANEA EMS 2017, a partir de la exposición de los objetivos, alcances y limitaciones de los resultados, para garantizar la validez de sus interpretaciones, su imparcialidad y señalar que se trata de información oportuna y adecuada.

Dentro de esta etapa, tiene una importancia fundamental el trabajo experto del Comité de Niveles de Logro y Puntos de Corte, que fue el cuerpo colegiado encargado establecer los niveles de desempeño, los cuales contemplaron la descripción de los niveles de competencia requeridos para responder los reactivos correctamente, y los puntos de corte, que fueron los indicadores que marcaron el límite entre un nivel y otro.

6.1. Comité de Niveles de Logro y Puntos de Corte

La evaluación basada en estándares determina los conocimientos, habilidades, objetivos o contenidos que se espera que un estudiante domine. Por ello, la determinación de niveles diferenciales de logro permite la interpretación cualitativa de estas habilidades de acuerdo con el puntaje que los estudiantes obtienen en la prueba. Cada nivel de logro es una definición operacional que representa un nivel de ejecución, el cual se realiza de manera acumulativa. Es a partir de las puntuaciones de corte (punto de inflexión entre dos niveles de logro) que se establecen las tipologías diferenciales de ejecución a lo largo del continuo.

El Comité de Niveles de Logro y Puntos de Corte contó con la participación de 28 docentes (14 por cada área disciplinar) que provenían de los diferentes subsistemas de la educación media superior (EMS). Estos especialistas trabajaron de manera colegiada en dos grupos:

- Grupo A: su primera tarea, previa al análisis empírico de resultados, fue identificar los descriptores genéricos de cada nivel de logro, en cada una de las áreas disciplinares evaluadas. Dichos descriptores fueron la guía para la identificación de las puntuaciones de corte que realizó el Grupo B. El segundo momento de su participación tuvo lugar con el ajuste de los descriptores considerando los resultados obtenidos por los estudiantes de la muestra y la identificación de las puntuaciones de corte definitivas.
- Grupo B: a partir de la revisión de los descriptores preparados por el Grupo A, la tarea central de este equipo fue identificar los reactivos marcadores y las puntuaciones de corte (la media del nivel de habilidad correspondiente a los reactivos identificados como marcadores), entre los diferentes niveles de logro y para cada área disciplinar.

En la siguiente tabla se muestran las instituciones que participaron en este Comité, agrupadas cada una a partir del área disciplinar correspondiente y los dos grupos de trabajo determinados para el desarrollo del Comité:

Tabla 25. Instituciones participantes en el Comité Niveles de Logro y Puntos de Corte	
Lenguaje y Comunicación	
Grupo A	**Grupo B**
Subsistema	**Subsistema**
UPN Ciudad de México	Particular Ciudad de México
UNAM Ciudad de México	CONALEP Ciudad de México
Telebachillerato Estado de México	UAEM
Telebachillerato Guanajuato	DGB CEB 4/1 Ciudad de México
COLBACH Ciudad de México	CEB 5/4. Pachuca de Soto Hidalgo
Particular	COLBACH Plantel 15, Cd México
COLBACH	CONALEP Estado de México
Matemáticas	
Grupo A	**Grupo B**
Subsistema	**Subsistema**
UAEM Estado de México	Telebachillerato Veracruz
COLBACH	CEB DGB Aguascalientes
Universidad de Yucatán	Preparatoria Universidad Autónoma de Guerrero
UNAM Ciudad de México	Liceo Alberto del Canto A. C. Nivel Bachillerato. Particular. Coahuila
Telebachillerato Guanajuato	Preparatoria Universidad Autónoma de Zacatecas
COLBACH Ciudad de México	CEB 6/9, Estado de México Ixtlahuaca
CEB DGB Ciudad de México	CONALEP Reynosa, Tamaulipas

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

6.1.1. Procedimiento metodológico

El procedimiento metodológico que guió las actividades del Comité de Niveles de Logro y Puntos de Corte se constituyó, como ya se mencionó, por dos grupos de trabajo, que tuvieron participación en diferentes momentos a lo largo del desarrollo del Comité, como se expone en la tabla 26, que expresa la organización general del Comité de Niveles de Logro y Puntos de Corte:

Tabla 26. Programación del Comité de Niveles de Logro y Puntos de Corte					
Día	**Lunes**	**Martes**	**Miércoles**	**Jueves**	**Viernes**
Grupo	A	B	B	B	A
	Niveles de Logro	Puntuaciones de Corte	Puntuaciones de Corte	Puntuaciones de Corte	Niveles de Logro

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

El primer momento del desarrollo metodológico del Comité fue el de la introducción a las características técnicas, metodológicas y conceptuales del diseño de PLANEA EMS 2017. Los jefes de prueba y la subdirectora presentaron a cada grupo de trabajo las atribuciones del INEE, los propósitos generales de PLANEA, las fases en el diseño de su construcción y la metodología utilizada para su concreción. Posteriormente, a cada grupo de trabajo se le presentaron sus objetivos particulares y la explicación del procedimiento para alcanzarlos.

Objetivos del Grupo A:

1. Analizar los indicadores de aprendizajes clave que se evaluaron en la aplicación de la prueba PLANEA EMS 2017 y clasificarlos en los cuatro niveles de logro definidos en el Plan Nacional para la Evaluación de los Aprendizajes.
2. Describir los cuatro niveles de logro a partir del análisis de las demandas cognitivas de los indicadores de aprendizaje. Insumo de trabajo para los integrantes del Grupo B del Comité Niveles de Logro.
3. Ajustar los descriptores a partir de las puntuaciones de corte identificadas por el Grupo B.
4. Elegir dos reactivos que ejemplifiquen cada nivel de logro (ocho reactivos en total).

Entre las especificaciones para este equipo de trabajo, se puntualizó que las descripciones debían señalar los conocimientos y habilidades que se expresaron a través de los reactivos de la prueba, pero éstas no debían referirse a contenidos o temas explícitos de los exámenes, sino a lo que los estudiantes podían realizar con base en éstos.

DÍA 1 (previo al análisis empírico de resultados por el Grupo B):

Los jefes de prueba y la subdirectora proporcionaron los insumos necesarios para que este primer grupo colegiado avanzara en sus tareas específicas; este material constó del listado de los indicadores de aprendizaje de los 148 reactivos que conformaron la prueba (de cada campo disciplinar y organizados por temática general), la definición de las categorías genéricas de los niveles de logro y las referencias de elaboración de descriptores de niveles de logro. Debe tomarse en cuenta que los descriptores genéricos de los niveles de desempeño que se utilizaron orientaron el trabajo del Comité en el desarrollo de los descriptores específicos de cada instrumento.

El Comité se desarrolló mediante sesión plenaria, la cual estuvo organizada, a su vez, en tres equipos de trabajo, quienes analizaron y discutieron la clasificación de los indicadores de aprendizaje en los cuatro niveles de logro. Cada equipo proyectó su propuesta de clasificación para cada una de las temáticas generales, la cual fue retroalimentada por los demás grupos, con la finalidad de llegar a una propuesta definitiva.

En un segundo momento, el cuerpo colegiado, nuevamente en plenaria, redactó la descripción de los cuatro niveles de logro para cada temática general, a partir de la clasificación de los indicadores de aprendizaje, y refirió las habilidades que se derivaron del logro de los mismos. Finalmente, se revisó el documento generado y se llegó a un acuerdo respecto a las descripciones de los cuatro niveles de logro.

DÍA 2 (posterior a la determinación de los Puntos de Corte):

Los jefes de prueba y la subdirectora presentaron a los expertos participantes los objetivos y la explicación del procedimiento para alcanzar el objetivo del Comité, además de proporcionar los insumos necesarios para el inicio de las tareas: a) las puntuaciones de corte que el Grupo B determinó a partir de los resultados obtenidos por los estudiantes, b) el documento con la descripción de los cuatro niveles de logro, c) el listado con los indicadores de aprendizaje y d) el cuadernillo con los 150 reactivos de la prueba PLANEA EMS 2017 para cada área disciplinar.

Para llevar a cabo el ajuste de los descriptores a partir de las puntuaciones de corte identificadas por el Grupo B, mediante plenaria, los integrantes del Grupo A analizaron estos resultados junto con las implicaciones de los resultados en las descripciones de los niveles de logro. A partir de este procedimiento, por equipos de trabajo se hicieron los siguientes ajustes a la redacción de la descripción de los cuatro niveles de logro de cada temática.

Posteriormente, el Comité analizó la coherencia de la descripción de los niveles de logro definitivos, eligió dos reactivos ejemplo de la dificultad media de los niveles de logro y, finalmente, revisó la pertinencia de los reactivos elegidos por el Grupo B, con lo cual dio término a sus actividades, mediante la firma del documento de la de descripción de los niveles de logro.

Objetivos del Grupo B:

1. Identificar los reactivos marcadores que operan como punto de inflexión entre dos niveles para determinar los niveles de logro de los resultados de la aplicación nacional de la prueba PLANEA EMS 2017, mediante el Método Bookmark.
2. Elegir dos reactivos que ejemplifiquen cada nivel de logro (ocho reactivos en total).

Día 1 (a partir de la revisión de los descriptores preparados por el Grupo A):

Los jefes de prueba y la subdirectora presentaron los objetivos específicos del Grupo de trabajo B y la explicación del procedimiento para alcanzarlos; proporcionaron además los insumos necesarios para que este segundo grupo colegiado avanzara en sus tareas específicas; este material constó de la definición de las categorías genéricas de los niveles de logro, la presentación de la descripción de los niveles de logro propuestos por el Grupo A y el cuadernillo de los reactivos que conformaron la prueba.

El inicio de las tareas se abrió con la resolución de todos los reactivos que conformaron la prueba PLANEA EMS 2017, con la finalidad de que el grupo de expertos analizara y reflexionara sobre las habilidades y conocimientos que exigían los ítems para ser correctamente contestados, además de que identificara cuáles eran los que correspondían a las habilidades que se describían en los cuatro niveles de logro.

Día 2 (a partir de la revisión de los descriptores preparados por el Grupo A):

En el segundo día de actividades, se capacitó a los participantes en el Método Bookmark (que se describe en el siguiente apartado de este Manual), para determinar el reactivo marcador que separa el nivel de logro 2 del nivel de logro 3. Los expertos de cada disciplina tuvieron como guía los descriptores que el Grupo A estableció en el día 1 de su trabajo y una pregunta guía: ¿Un estudiante del nivel 2 puede responder correctamente este reactivo? En plenaria, se llevaron a cabo tres rondas como mínimo y cinco como máximo.

Posteriormente, los docentes participantes tuvieron como tarea establecer el punto de corte 1 mediante el jueceo del Método Bookmark, con lo cual se esperó que determinaran el reactivo marcador que separa el nivel de logro 1 del nivel de logro 2, teniendo nuevamente la guía de los descriptores que el Grupo A desarrolló y la pregunta: ¿Un estudiante del nivel 1 puede responder correctamente este reactivo? En plenaria, se llevaron a cabo tres rondas como mínimo y cinco como máximo.

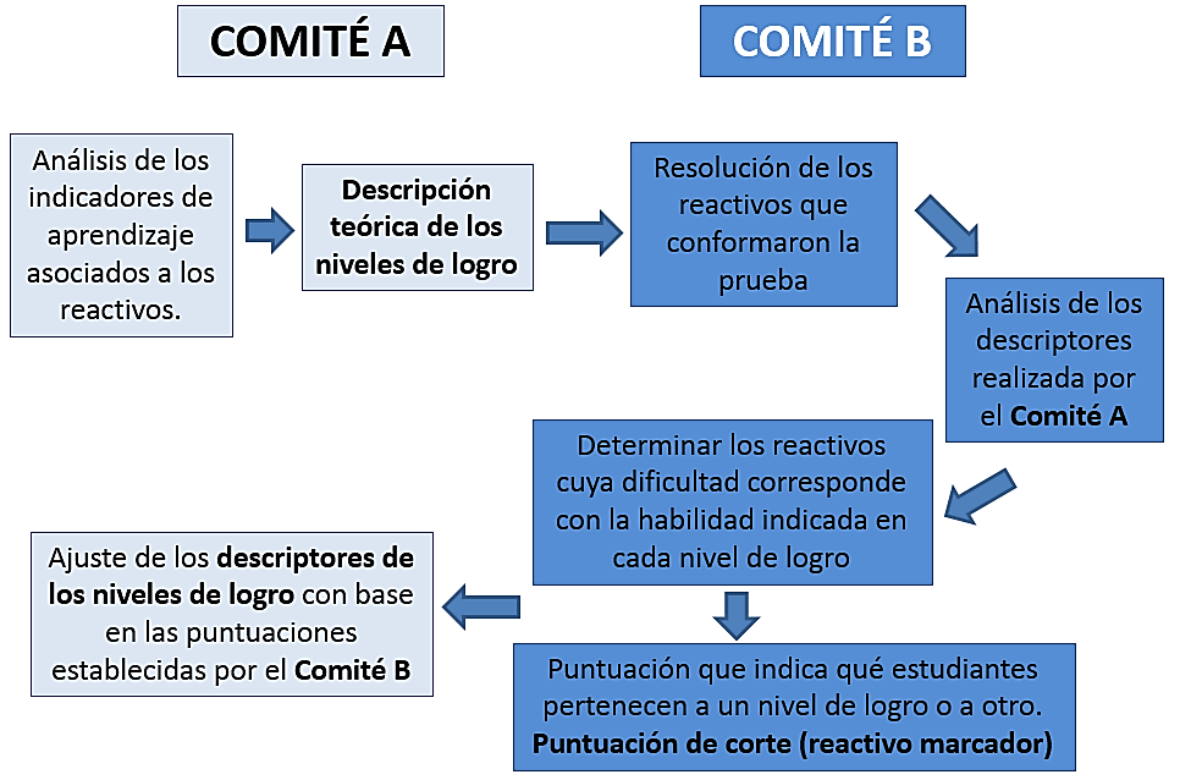
Día 3 (a partir de la revisión de los descriptores preparados por el Grupo A):

El tercer día de actividades del Grupo B, los expertos de cada área disciplinar determinaron el punto de corte 3 mediante el jueceo del Método Bookmark y determinaron el reactivo marcador que separa el nivel de logro 3 del nivel de logro 4. Los profesores respondieron al cuestionamiento: ¿Un estudiante del nivel 3 puede responder correctamente este reactivo?, y tomaron en cuenta los descriptores que el Grupo A desarrolló el día 1. Se llevaron a cabo tres rondas como mínimo y cinco como máximo.

Posteriormente, el grupo de expertos eligió, por cada área disciplinar, dos reactivos representativos de los valores de dificultad media de cada nivel de logro y, finalmente, firmó el documento de determinación de los puntos de corte, con lo cual cerró las actividades.

En el diagrama 3 se ilustra el proceso integrado del Comité de Niveles de Logro y Puntos de Corte.

Diagrama 3. Proceso integrado del Comité de Niveles de Logro y Puntos de Corte



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

6.1.2. Capacitación de los integrantes en el Método Bookmark

El método Bookmark fue el procedimiento para determinar las puntuaciones de corte que identificaron los niveles de logro. Este procedimiento consistió en señalar los intervalos que permitían decidir a qué nivel de logro pertenecían los estudiantes, a partir de la revisión de los ítems, presentados por gradaciones de dificultad (de acuerdo con la habilidad requerida para contestarlos correctamente). Es importante notar que la revisión de los reactivos permitió que los integrantes

del Comité tuvieran una visión global del grado de complejidad del instrumento de evaluación y que, con ello, pudieran centrar su atención en los ítems que constituyeron el límite entre un nivel y otro. Los descriptores genéricos de los niveles de logro apoyaron este proceso, mediante un cuestionamiento guía al que el grupo de expertos respondió para cada punto de corte: ¿Los estudiantes del nivel _____, con un logro _____ de los aprendizajes, pueden responder correctamente este reactivo?

A partir de la pregunta anterior, los expertos en el tema fueron discerniendo y colocando las marcas en los reactivos que consideraron límites, para posteriormente observar las discrepancias que fueron surgiendo entre los reactivos marcadores señalados por cada juez. Con esto fue necesario dar paso a una discusión detallada sobre lo que el reactivo medía en sí mismo y por qué mostraba una mayor complejidad que el ítem anterior.

En el diagrama 4 se detallan los pasos que los integrantes del Comité siguieron para efectuar de manera adecuada esta tarea:

Diagrama 4. Pasos para establecer las puntuaciones de corte

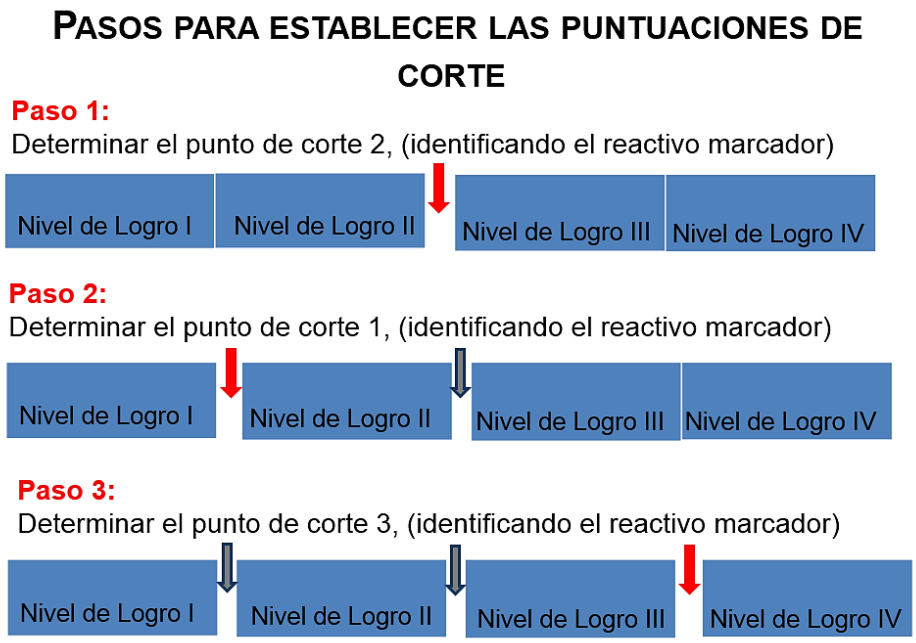
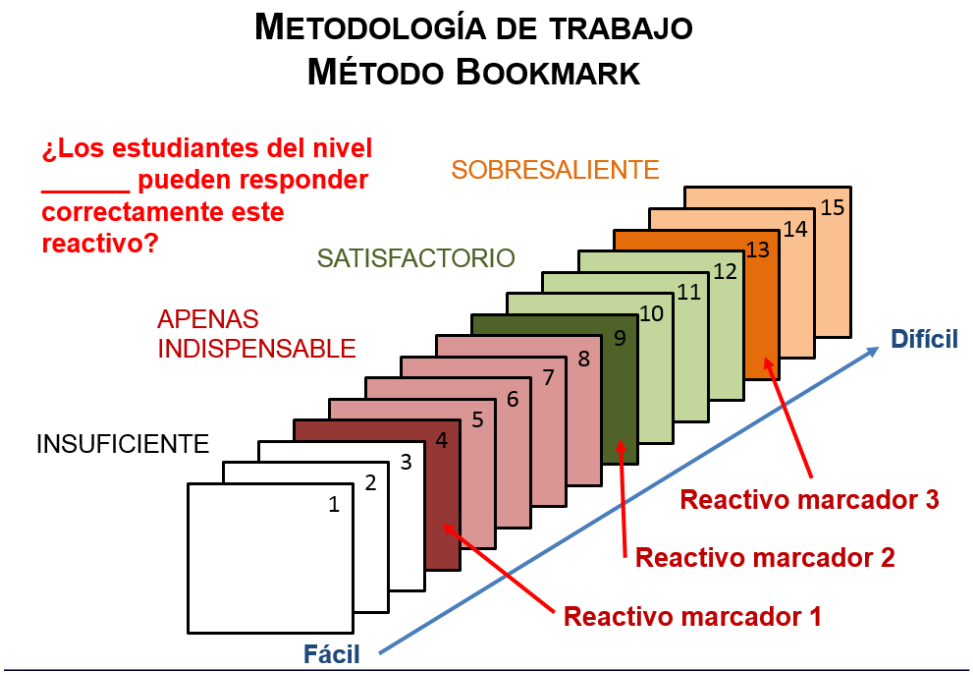


Diagrama 5. Método Bookmark



Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

6.2. Resultados

En las tablas 28 a 34 se presentan los resultados del trabajo del Comité de Niveles de Logro y Puntos de Corte.

- **Niveles de Logro Lenguaje y Comunicación (Método Bookmark)**

Tabla 27. Descripción de los niveles de logro de Lenguaje y Comunicación (Método Bookmark)			
Manejo y construcción de la información			
Nivel I	Nivel II	Nivel III	Nivel IV
• Seleccionan información para hacer un trabajo escrito. • Diferencian entre una encuesta, una entrevista y el cuestionario. Sin embargo, tienen dificultad para plantear problemas, clasificar información de fuentes de consulta, parafrasear información e interpretar textos.	• Identifican las etapas del proceso de redacción. • Identifican la coherencia y la cohesión de un texto. • Distinguen las características de los trabajos académicos. • Seleccionan fuentes de consulta adecuadas para el tema de interés. • Identifican los diferentes tipos de ficha. • Reconocen el proceso de investigación. • Planean la estructura de un texto de acuerdo con un propósito comunicativo.	• Identifican las funciones comunicativas de la lengua. • Implementan estrategias de comprensión lectora. • Organizan las actividades de un proyecto de trabajo. • Resumen textos. • Distinguen las características de distintos tipos de texto. • Aplican términos especializados de acuerdo con el contexto. • Reconocen prejuicios y valores éticos en un texto.	• Plantean problemas para una investigación. • Clasifican información de diferentes fuentes de consulta. • Conocen los conceptos de preescritura y borrador. • Infieren la paráfrasis de un texto expositivo como artículos de divulgación. • Interpretan textos de acuerdo al contexto de recepción y al de producción.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 28. Descripción de los niveles de logro de Lenguaje y Comunicación (Método Bookmark)			
Texto argumentativo			
Nivel I	Nivel II	Nivel III	Nivel IV
• Identifican el inicio, desarrollo y conclusión de un texto argumentativo. Sin embargo, tienen dificultad para organizar esquemas argumentativos, identificar problemas sociales, económicos o políticos en un escrito, identificar la postura del autor en una reseña crítica y establecer diferencias entre textos demostrativos y textos persuasivos.	• Distinguen el propósito comunicativo de textos periodísticos. • Identifican los argumentos de un texto. • Seleccionan argumentos para un ensayo. • Sintetizan un texto argumentativo.	• Identifican el concepto de argumentación. • Reconocen los elementos de los textos argumentativos. • Seleccionan los recursos textuales y las formas lingüísticas propias de la argumentación. • Distinguen un argumento válido de una falacia.	• Organizan esquemas argumentativos. • Identifican problemas sociales, culturales, económicos o políticos en un texto. • Diferencian entre un texto demostrativo y uno persuasivo. • Identifican una reseña crítica. • Seleccionan los argumentos para una reseña crítica. • Identifican la postura del autor de una reseña crítica.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 29. Descripción de los niveles de logro de Lenguaje y Comunicación (Método Bookmark)			
Texto expositivo			
Nivel I	Nivel II	Nivel III	Nivel IV
• Identifican el tema central, el problema planteado y la estructura de un texto expositivo. Sin embargo, tienen dificultad para identificar el concepto de referente, deducir el significado de palabras y realizar una lectura crítica de estos textos.	• Identifican el significado de los textos expositivos. • Diferencian los propósitos comunicativos de este tipo texto y otros. • Deducen el significado de las palabras dentro de un texto.	• Identifican el concepto de referente.	• Realizan la lectura crítica de textos expositivos.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 30. Descripción de los niveles de logro de Lenguaje y Comunicación (Método Bookmark)			
Texto literario			
Nivel I	Nivel II	Nivel III	Nivel IV
• Distinguen la información explícita dentro de una secuencia narrativa. Sin embargo, tienen dificultad para deducir la secuencia narrativa de una obra de teatro, identificar las características de los textos literarios, distinguir la información implícita en una secuencia narrativa compleja y analizar el contexto de una novela, una obra de teatro o un poema.	• Identifican las características de un texto narrativo sencillo. • Infieren la secuencia narrativa de una obra de teatro.	• Identifican el propósito comunicativo del texto literario. • Comparan textos literarios con los de medios de comunicación.	• Identifican las características de los textos literarios. • Distinguen la información implícita de una secuencia narrativa compleja. • Analizan el contexto de una novela, una obra de teatro o un poema.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

- **Niveles de Logro Matemáticas (Método Bookmark)**

Tabla 31. Descripción de los niveles de logro de Matemáticas (Método Bookmark)			
Sentido numérico y pensamiento algebraico			
Nivel I	Nivel II	Nivel III	Nivel IV
• Los estudiantes resuelven problemas que implican operaciones básicas con números enteros o cuyo resultado es un número entero. Sin embargo, tienen dificultades para emplear algoritmos aritméticos más elaborados y dificultades importantes en el dominio del álgebra.	• Resuelven problemas aditivos con fracciones con denominador común, y que implican el cálculo directo de razones o porcentajes.	• Resuelven operaciones que involucran números enteros y signos de agrupación. • Resuelven multiplicaciones de polinomios. • Resuelven problemas aditivos de fracciones y que implican el planteamiento de ecuaciones.	• Resuelven operaciones que involucran números reales y signos de agrupación. • Resuelven problemas multiplicativos de fracciones mixtas. • Resuelven restas de polinomios y divisiones de polinomios entre monomios.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 32. Descripción de los niveles de logro de Matemáticas (Método Bookmark)			
Cambios y relaciones			
Nivel I	Nivel II	Nivel III	Nivel IV
• Los estudiantes resuelven problemas de valor faltante en tablas de proporcionalidad directa. • Identifican el valor máximo que alcanza un fenómeno a partir de su gráfica. Sin embargo, tienen dificultades para reconocer y establecer, algebraica o gráficamente, la relación de dependencia de dos variables.	• Resuelven problemas de valor faltante en tablas de proporcionalidad inversa. • Identifican la función lineal que modela a un fenómeno	• Determinan si los datos de una tabla presentan relaciones de proporcionalidad. • Resuelven problemas de proporcionalidad. • Interpretan las relaciones y parámetros de la función lineal dentro una situación. • Resuelven la suma de funciones y evalúan números positivos en ellas.	• Resuelven la resta de funciones y evalúan números negativos en ellas. • Determinan el dominio y el rango de una función, así como el valor de la pendiente y la ecuación de una recta a partir de su gráfica.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 33. Descripción de los niveles de logro de Matemáticas (Método Bookmark)			
Forma, espacio y medida			
Nivel I	Nivel II	Nivel III	Nivel IV
• Los estudiantes tienen dificultades para aplicar las propiedades de las figuras geométricas para resolver problemas.	• Resuelven problemas que implican el cálculo de ángulos en intersección de rectas por medio de la aplicación directa de una sola propiedad.	• Resuelven problemas que implican el cálculo de ángulos en intersección de rectas por medio de la aplicación de más de una propiedad. • Determinan por qué dos triángulos son semejantes.	• Resuelven problemas aplicando el teorema de Pitágoras para calcular uno de los catetos o determinar si un triángulo es un triángulo rectángulo.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 34. Descripción de los niveles de logro de Matemáticas (Método Bookmark)			
Manejo de la información			
Nivel I	Nivel II	Nivel III	Nivel IV
• Los estudiantes resuelven problemas de moda y media aritmética para datos enlistados. • Interpretan la posibilidad de ocurrencia de los eventos de un experimento a partir de una gráfica de frecuencias. Sin embargo, tienen dificultades para establecer las medidas de tendencia central cuando los datos no se presentan enlistados o para calcular probabilidades.	• Resuelven problemas de moda cuando los datos se organizan en varias tablas. • Determinan el valor del dato faltante de un conjunto para ajustar su media aritmética a un valor preestablecido. • Calculan la probabilidad de un evento simple.	• Interpretan y abstraen información que se presenta en gráficas. • Calculan la probabilidad de eventos compuestos. • Comparan e interpretan las probabilidades asociadas a los eventos de una experiencia aleatoria.	• Determinan la mediana de un conjunto de datos, para un número par de ellos. • Resuelven problemas de media aritmética cuando los datos se presentan en histogramas.

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

6.3. Análisis estadístico de los resultados de la prueba nacional (considerando los puntos de corte de los niveles de logro determinados en el Comité)

A continuación se muestra el porcentaje de estudiantes en cada nivel de logro a nivel nacional. El análisis completo, que contiene los puntajes por tipo de control administrativo, por tipo de servicio, por capital económico, por sexo, por edad, por condición de habla de lengua indígena y por nivel de estudios de los padres, puede ser consultado en línea en: http://publicaciones.inee.edu.mx/buscadorPub/P2/A/328/P2A328.pdf.

Tabla 35. Porcentaje de estudiantes en cada nivel de logro Lenguaje y Comunicación		
Nivel IV	Seleccionan y organizan información pertinente de un texto argumentativo; identifican la postura del autor, interpretan información de textos argumentativos (como reseñas críticas y artículos de opinión) e infieren la paráfrasis de un texto expositivo (como un artículo de divulgación).	9.2
Nivel III	Reconocen en un artículo de opinión: propósito, conectores argumentativos y partes que lo constituyen (tesis, argumentos y conclusión); identifican las diferencias entre información objetiva, opinión y valoración del autor; identifican las diferentes formas en que se emplea el lenguaje escrito de acuerdo con la finalidad comunicativa y utilizan estrategias para comprender lo que leen.	28.7
Nivel II	Identifican ideas principales que sustentan la propuesta de un artículo de opinión breve, discriminan y relacionan información oportuna y confiable, y la organizan a partir de un propósito.	28.1
Nivel I	No identifican la postura del autor en artículos de opinión, ensayos o reseñas críticas; ni explican la información de un texto sencillo con palabras diferentes a las de la lectura.	33.9

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

Tabla 36. Porcentaje de estudiantes en cada nivel de logro Matemáticas		
Nivel IV	Dominan las reglas para transformar y operar con el lenguaje matemático (por ejemplo, las leyes de los signos); expresan en lenguaje matemático las relaciones que existen entre dos variables de una situación o fenómeno; y determinan algunas de sus características (por ejemplo, deducen la ecuación de la línea recta a partir de su gráfica).	2.5
Nivel III	Emplean el lenguaje matemático para resolver problemas que requieren del cálculo de valores desconocidos, y para analizar situaciones de proporcionalidad.	8.0
Nivel II	Expresan en lenguaje matemático situaciones donde se desconoce un valor o las relaciones de proporcionalidad entre dos variables, y resuelven problemas que implican proporciones entre cantidades (por ejemplo, el cálculo de porcentajes).	23.3
Nivel I	Tienen dificultades para realizar operaciones con fracciones y operaciones que combinen incógnitas o variables (representadas con letras), así como para establecer y analizar relaciones entre dos variables.	66.2

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS. INEE, México.

6.4. Interpretación del análisis estadístico de los resultados de la prueba nacional

A continuación se presenta la interpretación que el INEE desarrolló a partir de los resultados estadísticos de la prueba PLANEA EMS 2017.

- Los resultados de PLANEA ELSEN son una fotografía de los niveles de logro, en Lenguaje y Comunicación y Matemáticas, de los estudiantes que terminan la educación media superior (EMS).
- La diversidad de los tipos de servicio en la EMS es una consecuencia histórica y social del establecimiento del Sistema Educativo Nacional (SEN), lo que aunado a factores sociales, culturales y económicos, influye en los resultados de logro educativo. Por ello, la mejora del logro educativo exige una atención diferenciada en cada entidad, tipo de servicio y tipo de control administrativo.
- Los resultados presentados confirman los bajos niveles de desempeño de los alumnos en el SEN reportados en diversas evaluaciones del logro educativo, tanto nacionales como internacionales, a lo largo de la educación obligatoria. En Lenguaje y Comunicación, 34% de los estudiantes se ubicó en el nivel de logro I (el más bajo), y 66% en el nivel I de Matemáticas; estos estudiantes no han consolidado los aprendizajes clave que se evaluaron en la prueba PLANEA EMS 2017, tales como realizar inferencias de contenidos implícitos en diferentes tipos de texto o realizar inferencias a partir de un modelo matemático.
- A nivel nacional, en Lenguaje y Comunicación, 9% de los estudiantes se encuentra en el nivel de logro IV (el más alto), y en Matemáticas, 3 por ciento.
- Los resultados son un reflejo de múltiples factores, desde las actividades escolares de los estudiantes (hábitos, actitudes y valores), hasta las condiciones de las instituciones educativas y el contexto socioeconómico en el que viven, entre otros.
- Los resultados por tipo de control administrativo y por tipo de servicio pueden estar relacionados en cierta medida con la demanda y los procesos de selección de los bachilleratos autónomos, así como con el prestigio que éstos tienen al depender de las universidades públicas autónomas. Además, es posible que egresar de este tipo de bachillerato otorgue a los estudiantes mejores oportunidades para ingresar a la universidad.
- El logro académico de los estudiantes de EMS está vinculado con los resultados de los niveles educativos previos. Para atender esta situación, es necesario reducir las brechas de conocimiento, oportunidades y condiciones generales de la enseñanza y del aprendizaje, desde el inicio de la educación obligatoria.
- Si bien es evidente que un considerable porcentaje de estudiantes se ubicó en el nivel de logro más bajo, sería deseable que la mayoría de ellos, en ambas asignaturas, alcanzaran no sólo el nivel II sino al menos el nivel III, pues esto indicaría que tienen un dominio satisfactorio de los aprendizajes clave. No obstante, en Lenguaje y Comunicación observamos que para 1 de cada 3 estudiantes, continuar con sus estudios profesionales puede significar un mayor esfuerzo que para el resto de sus compañeros. Esta situación se acentúa en Matemáticas, donde 2 de cada 3 estudiantes se encuentran en dicha situación. Los resultados mostrados [...] representan un enorme reto para la educación media superior e implica que también se requieren esfuerzos coordinados de muchos actores de niveles educativos previos para que todos los alumnos puedan ejercer plenamente su derecho a recibir una educación de calidad.

- Es de suma importancia analizar la inequidad en los resultados educativos. Una forma para observarla es cuando hay diferencias entre los puntajes que obtienen los alumnos que asisten a distintos tipos de servicio. Idealmente, sin importar las condiciones de origen y la escuela a la que asistan, todos los estudiantes deberían dominar un conjunto de contenidos curriculares básicos. Los resultados en EMS evidencian la inequidad que existe entre los estudiantes que asisten a los distintos tipos de bachillerato, por ejemplo, se observa una gran distancia entre los resultados de los telebachilleratos con respecto a los particulares y autónomos. Si se da un seguimiento periódico, será posible saber si las distancias se acortan. Por ello, la distribución de recursos destinados a la educación debería dar prioridad a las poblaciones cuyos tipos de servicio se encuentran en desventaja para lograr mejores resultados de aprendizaje y un progreso que permita disminuir las brechas educativas.
- Se espera que quienes concluyan sus estudios en la EMS hayan desarrollado la capacidad para pensar, reflexionar, analizar y comunicarse, así como para emplear y transformar los aprendizajes matemáticos en herramientas que les permitan solucionar diferentes problemas. Estas habilidades son importantes, tanto para aquellos que continúen con sus estudios, como para quienes se incorporen al campo laboral.
- El análisis detallado de los niveles de logro permite conocer los aciertos y retos que se tienen en el aprendizaje de los contenidos de las áreas evaluadas (INEE, 2017b: s.n.p.).

6.4.1. Reporte de resultados

Como se ha mencionado, las evaluaciones a gran escala favorecen la toma de decisiones para el mejoramiento de la educación de calidad y con equidad. Por ello, la difusión de los resultados que éstas generan debe realizarse de manera amplia y oportuna entre los diversos actores que tienen incidencia en las políticas educativas, así como entre la sociedad civil. PLANEA, en su documento rector expresa lo siguiente al respecto:

> Todo sistema de evaluación se funda en la premisa de que la información y el conocimiento que genera deberán ser utilizados para introducir mejoras en el objeto evaluado. Por ello, una de las preocupaciones fundamentales que acompañan la devolución de resultados de[l] PLANEA es la pertinencia de la información que se ofrece, y la identificación de las necesidades de información de distintos tipos de usuarios del sistema educativo. Se espera que los usuarios aprovechen la información producida por la evaluación para enriquecer sus decisiones y de esta manera contribuir a la mejora educativa [desde la promoción y capacitación en el uso adecuado de resultados] (INEE, 2018: 22).

Así, y con respecto a la difusión de resultados, el INEE genera dos tipos de reportes sobre los resultados del PLANEA:

1. Reporte general de resultados de[l] PLANEA SEN. Documento breve, descriptivo, que incluye información nacional, estatal y de los dominios de estudio principales, con apoyos gráficos y textos breves. Se presenta cinco meses después de la aplicación a las autoridades educativas federal y estatales, y sus resultados son comentados en mesas públicas organizadas por el Instituto. Además de información sobre el dominio de diferentes grupos de estudiantes de los aprendizajes clave, el reporte debe incluir información específica que muestre las diferencias en los resultados entre los grupos sociales más favorecidos y los más desfavorecidos, de manera que sean muy claras las brechas que hay que eliminar para alcanzar la equidad educativa.

2. Bases de datos. El INEE pone a disposición de autoridades educativas, universidades, organizaciones de la sociedad civil, medios de comunicación y público en general, al mismo tiempo que el Informe nacional, de manera que sea posible, a los usuarios especializados, realizar análisis complementarios a los que lleva a cabo el INEE. Además, desarrollará componentes temáticos en el Sistema Integral de Resultados de las Evaluaciones (SIRE) que permitan explorar de forma dinámica la relación entre los resultados de PLANEA SEN y diferentes elementos de contexto, con la intención de apoyar los análisis que desarrollen autoridades y actores educativos, investigadores, así como diversos usuarios de la sociedad civil interesados en el tema educativo (INEE, 2018: 24).

Adicionalmente, el INEE también podrá elaborar:

3. Materiales de divulgación. El INEE podrá desarrollar diferentes materiales como fascículos para docentes, videos, podcasts, infografías, trípticos para la divulgación con docentes, padres de familia, organizaciones de la sociedad civil, etc., respecto de algunos de los resultados de PLANEA que sean de particular interés para dichos grupos.
4. Informes temáticos. El Instituto podrá desarrollar informes centrados en alguna temática en particular, apoyados en los resultados de las evaluaciones de PLANEA SEN, por ejemplo, sobre los resultados de los estudiantes en condiciones de extraedad con respecto a los estudiantes con un trayecto escolar regular, sobre los resultados en las escuelas multigrado, o sobre la identificación de características escolares y familiares de grupos con menores niveles de logro (INEE, 2018: 24).

- **Reporte de resultados en los estados**

Las Direcciones del INEE en las entidades federativas (DINEE) también presentaron resultados por cada entidad. Éstos se generan a partir de una comparativa con el puntaje de la media nacional. Así, se obtiene de manera estadística cuál es el logro alcanzado por los estudiantes de cada entidad con respecto al promedio nacional. Las DINEE también generan datos a partir del porcentaje de sus alumnos en cada uno de los niveles de logro. Esta información permite realizar una estimación de cuáles son las entidades federativas con mayor y menor porcentaje con respecto a la media nacional, así como la diferencia entre el puntaje mayor y el menor. Dichos elementos tienen la finalidad de evidenciar las problemáticas a la que se enfrenta cada DINEE con respecto al logro educativo de sus alumnos.

6.4.2. Informe institucional

El informe de PLANEA SEN es un documento extenso en el que se informa a las autoridades educativas y la sociedad civil sobre el estado que guarda la educación en México:

> con respecto al logro de aprendizajes a partir de los resultados de cada aplicación, que puede verse enriquecido con datos de aplicaciones previas [...]. Se entrega 14 meses después de la aplicación. Profundiza en los resultados entregados en el Reporte de[l] PLANEA SEN, y agrega información sobre la relación de variables de contexto con el aprendizaje (INEE, 2017b: s.n.p.).

En el caso de la EMS, el informe institucional puede ser consultado en la siguiente liga: http://publicaciones.inee.edu.mx/buscadorPub/P2/A/328/P2A328.pdf.

6.4.3. Cuestionarios de contexto

Para la interpretación de resultados, además de los análisis mencionados, PLANEA EMS 2017 también consideró los datos de los cuestionarios de contexto aplicados a los alumnos, ya que el elemento contextual del desarrollo del estudiante impacta de manera importante en el logro educativo de éste. Así pues, el documento rector de PLANEA desarrolla la importancia de este tipo de instrumentos de la siguiente manera:

> Es conocido que diversas circunstancias influyen en el logro de aprendizajes de los estudiantes. La investigación educativa señala que entre las condiciones personales y familiares que inciden en el aprendizaje se encuentran factores socioeconómicos, acceso a bienes culturales, condiciones para el estudio, situación laboral de la familia, pertenencia étnica, motivación personal, expectativas, disposición e interés por el estudio, trayectoria escolar, etc. Otros factores que influyen en el aprendizaje están relacionados con las características de la escuela (infraestructura, equipamiento y organización), o con los procesos pedagógicos, las trayectorias y las habilidades de enseñanza de los docentes, el liderazgo de los directivos y su capacidad de gestión, entre otros.
>
> Para contextualizar y comprender las diferencias en los resultados de aprendizaje, tanto al nivel de sistema educativo como de plantel escolar, es necesario considerar los factores personales, familiares, del entorno escolar y de las condiciones en las que trabajan las comunidades educativas, así como las características, los rasgos y las prácticas de quienes las componen. Poder establecer los factores internos y externos que se asocian con la promoción de los aprendizajes ayuda a identificar aquello sobre lo que es posible incidir para mejorar las oportunidades de todos los alumnos e incrementar la calidad de la educación.
>
> Por esta razón, junto con los instrumentos de evaluación, se aplican a alumnos, docentes, directivos y padres de familia instrumentos denominados cuestionarios de contexto que permiten obtener información sobre condiciones personales, familiares y escolares, como etnicidad, nivel de estudios de los padres, recursos familiares asociados al bienestar, oportunidades para el aprendizaje dentro y fuera de la escuela, trayectoria escolar, formación y trayectoria profesional de docentes y directivos, organización escolar, infraestructura de la escuela, clima escolar, liderazgo del director, entre otras.
>
> Estos cuestionarios, al igual que las pruebas, se revisan de manera periódica, con el fin de realizar los ajustes necesarios, en particular para la incorporación de temas emergentes que se juzgue necesario tomar en cuenta, así como mejorar las escalas existentes en función de su comportamiento psicométrico en las distintas aplicaciones. No obstante, se busca conservar una base uniforme de información de los llamados factores de contexto entre aplicaciones, que ayude a comprender los resultados de logro de aprendizajes y sus cambios a lo largo del tiempo (INEE, 2018: 18).

En el caso de PLANEA EMS 2017, los Cuestionarios para el alumno se aplicaron en la jornada 1, después de la prueba de Lenguaje y Comunicación. Dichos cuestionarios contuvieron un total de 120 reactivos que permitieron recabar las características personales, familiares, escolares y sociales de los estudiantes a quienes se evaluó mediante PLANEA. Entre los índices contemplados para PLANEA EMS 2017 se encontraron los siguientes:

1. Edad y sexo del alumno: el cálculo de la edad y sexo del alumno implicó un proceso minucioso derivado de considerar tres variables provenientes de dos fuentes de información: a) la CURP proveniente de la información oficial [...] muestreo y b) los años cumplidos al momento de la prueba y el mes de nacimiento provenientes de las respuestas del alumno.
2. Condición indígena: este índice reporta la respuesta del alumno a la pregunta del Cuestionario de contexto "¿Alguno de tus padres habla una lengua indígena?". Se trata de un índice compuesto de dos categorías válidas: 1: Sí y 2: No. De modo que, cuando el alumno reporta que al menos uno de sus padres habla una lengua indígena entonces "Condición indígena" vale 1, si el alumno reporta que tanto su madre como su padre no hablan una lengua indígena entonces "Condición indígena" vale 2, si ninguna de estas condiciones se cumple entonces "Condición indígena" es un valor perdido (9).
3. Nivel de estudios del padre y nivel de estudios de la madre: estos dos índices reportan la respuesta del alumno a la pregunta del Cuestionario de contexto "¿Cuál es el nivel de estudios alcanzado por tu padre/madre (contesta aunque ya haya fallecido)?".
4. Máximo nivel de estudios alcanzado por los padres: para construir este índice se obtuvo la respuesta de los alumnos a la pregunta: "¿Cuál es el nivel de estudios alcanzado por tu padre (contesta aunque ya haya fallecido)?".
5. Capital económico: estos reactivos hacen alusión a factores materiales como el número de aparatos en casa, y los bienes y servicios de que dispone el alumno en su hogar (DGMTD, 2018b: 5-10).

Para más información acerca de los índices simples y las escalas de contexto, ver el anexo XIV.

6.4.4. Usos de los resultados

Una de las etapas finales es la difusión de resultados. Sin esta fase, el proceso de evaluación estaría incompleto, ya que la difusión y uso de resultados permite llevar a cabo acciones concretas en la toma de decisiones en política educativa, las cuales contribuyan a la promoción de una cultura de la evaluación.

A continuación se presentan algunos usos de resultados presentados por PLANEA EMS 2017.

1. Como referentes para la toma de decisiones en el ámbito educativo, el monitoreo, la planeación, la programación y la operación del sistema educativo y de los centros escolares, la capacitación y actualización docente, el diseño de medidas específicas que permitan favorecer los aprendizajes.
2. En el ámbito de la responsabilidad municipal los resultados pueden ser útiles en la toma de decisiones sobre la distribución y el uso de los recursos, para así mejorar las condiciones de la oferta educativa, como infraestructura y servicios.
3. Los resultados divulgados en cualquier reporte o informe público, así como las respectivas bases de datos, se pueden utilizar para el desarrollo de trabajos de investigación en la materia.
4. La sociedad puede utilizar los resultados divulgados en cualquier reporte o informe público para conocer el estado general de la educación en el país en términos del logro de los aprendizajes y tomar decisiones informadas a fin de hacer exigible el derecho de niñas, niños y jóvenes a una educación de calidad (INEE, 2016b: 39-43).

7. Mantenimiento del instrumento de evaluación

La etapa del mantenimiento de un instrumento de evaluación resulta crucial si la prueba seguirá funcionando, ya que es el momento en que se valora la experiencia que se obtuvo, se hace una revisión de los análisis realizados y se puede establecer un plan de mejora o de fortalecimiento de los puntos positivos. Las revisiones de los métodos, instrumentos y estrategias de trabajo de PLANEA EMS 2017 contribuyen al fortalecimiento de la calidad técnica del instrumento de evaluación; además de que constituyen una obligación, por parte del Instituto Nacional para la Evaluación d ela Educación (INEE), para garantizar la vigencia de los horizontes técnicos, los marcos conceptuales y la metodología que guían el desarrollo de PLANEA.

Por esta razón, a continuación se expone un plan de trabajo para enriquecer la calidad de este instrumento de evaluación en subsecuentes aplicaciones.

7.1. Indicadores de validez

Los indicadores de validez están relacionados directamente con la revisión y la actualización del objeto de medida de la evaluación, ya que éstos arrojan información estadística que permite observar el comportamiento de los ítems que constituyen el instrumento. Así, los *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* señalan que entre los indicadores de validez a considerar están los análisis de dimensionalidad, de confiabilidad y del funcionamiento diferencial, pues éstos permiten "suponer un comportamiento errático del instrumento o en alguno de sus componentes o dimensiones. Esta información debe ser el fundamento para orientar la revisión y los posibles cambios para mejorar la precisión de las inferencias que se realizan" (INEE, 2017: 48).

7.1.1. Análisis de dimensionalidad, confiabilidad y funcionamiento diferencial del instrumento

Tanto el marco teórico de la medición como la metodología utilizada para llevar a cabo los análisis referidos a la aplicación nacional de PLANEA EMS 2017 fueron descritos ya en el apartado 7. Análisis de resultados del instrumento de evaluación. En este apartado sólo se presenta los resultados de dichos análisis, para ilustra la manera en que se comportaron los reactivos de PLANEA EMS 2017.

- **Dimensionalidad**

La dimensión se puede definir como un elemento resultante de la descomposición de una variable compleja. En el caso de PLANEA EMS 2017, resultó adecuado realizar un ajuste al modelo de Rasch, que consistió en el análisis a partir de dos dimensiones (los dos campos disciplinares evaluados). La intención fue comparar los modelos unidimensional y bidimensional, para observar el nivel de confiabilidad que cada una de las pruebas arrojaba según el modelo elegido. Tras correr el análisis comparando ambos modelos, se corroboró que la confiabilidad aumentó y los puntajes mostraron una menor dispersión, lo cual se explicó en el apartado 7.3.2. Reporte preliminar del análisis, de este Manual técnico.

De esta manera, el análisis bidimensional resultó adecuado para los fines de la prueba y, además, resultó útil para calcular las habilidades de los estudiantes evaluados con la metodología de EAP/PV, como a continuación se detalla.

- **Confiabilidad**

La confiabilidad es un indicador cuantitativo con el que se mide el grado en que un instrumento de evaluación educativa produce resultados consistentes y coherentes. En el caso de PLANEA EMS 2017, el análisis que se utilizó para medir la confiabilidad fue el modelo bidimensional de Rasch, y el índice utilizado fue el Expected A Posteriori/Plausible Value (EAP/PV), el cual es una estimación de la confiabilidad con que se pueden usar los elementos para distinguir las habilidades subyacentes de los estudiantes.

Los resultados reportados por la Dirección General de Medición y Tratamiento de Datos (DGMTD) después del análisis revelaron una confiabilidad satisfactoria de EAP/PV: 0.840 para la prueba de Lenguaje y Comunicación y 0.898 para Matemáticas.

La explicación del proceso metodológico seguido para llegar a estas conclusiones se puede leer en el apartado 7.3.2. Reporte preliminar del análisis.

- **Funcionamiento diferencial**

Como se explicó en el apartado 7, el funcionamiento diferencial (DIF) tiene lugar cuando un grupo focal tiene desventaja (menor probabilidad de contestar correctamente un reactivo) frente a un grupo de referencia, aunque ambos posean el mismo nivel de habilidad.

Los análisis realizados al PLANEA EMS 2017 mostraron que la gran mayoría de reactivos no tenía DIF, ya que, de los 148 ítems utilizados para la medición, sólo se eliminaron 2 por cada área disciplinar; es decir, 4 reactivos en total. Lo anterior demuestra que los ítems de la prueba fueron construidos con solidez técnica y equidad para la mayoría de los evaluados.

Para consultar el análisis del funcionamiento diferencial de los reactivos de la aplicación nacional, éstos también se pueden encontrar en el apartado 7.3.2. Reporte preliminar del análisis.

7.2. Construcción del banco de reactivos

El banco de reactivos es un:

> Repositorio donde se resguardan y clasifican los reactivos que integran los instrumentos de evaluación; en él se administran los datos de identificación del reactivo, sus características métricas, las formas en las que se incorporó y las fechas en las que se utilizó (INEE, 2017: 6).

La elaboración de este repositorio fue un paso necesario, ya que, al ser la primera vez que el INEE diseñó PLANEA SEN para la educación media superior (EMS), debió integrar el registro de las personas que intervinieron en la elaboración de los ítems y las cualidades métricas de los mismos. Este banco de reactivos facilita el acceso a información especializada, pero estableciendo protocolos donde se especifiquen las políticas y procedimientos para el resguardo de esta base de datos.

Además, el banco de reactivos no sólo hace referencia de la introducción de la información en un sistema, sino que también promueve la revisión de los ítems a partir de características cualitativas y cuantitativas. Esta revisión, que se debe realizar en cada nueva aplicación de la prueba, tiene la finalidad de determinar si se cuenta con ítems de calidad, con buenos indicadores psicométricos y,

sobre todo, si éstos permiten conformar nuevos cuadernillos de acuerdo con la estructura vigente de la prueba.

Respecto a lo anterior, es necesario recordar que PLANEA es un instrumento de evaluación que tiene revisiones periódicas para mejorar su calidad. Aunado a lo anterior, también se prevé:

> [...] una revisión profunda de sus alcances y limitaciones tras el primer ciclo completo de aplicación, durante 2021. Esta revisión deberá ser llevada a cabo por especialistas externos al INEE, a quienes se proporcionará toda la información pertinente para que puedan hacer recomendaciones que mejoren las prácticas de desarrollo, aplicación y análisis de las pruebas, y que potencien los usos de sus resultados. Con base en estas recomendaciones el Instituto podrá ratificar o redefinir los propósitos de[l] PLANEA, sus horizontes técnico y logístico, así como su programa de aplicaciones (INEE, 2018: 26).

Así pues, el banco de reactivos de EMS debe ser un repositorio en constante revisión adaptable a las nuevas orientaciones que se propongan para PLANEA.

7.3. Manual técnico del instrumento de evaluación

La elaboración de un Manual técnico en el que se reporten las evidencias de la construcción, administración y análisis de resultados del instrumento de evaluación constituye uno de los pasos que el inee, a través de sus *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, señala como necesario para documentar los procesos que se siguieron en el diseño de PLANEA EMS 2017.

Así pues, el presente Manual técnico no sólo describe los pasos metodológicos y el marco de referencia de la construcción de PLANEA EMS 2017, también tiene una intención explicativa, con la finalidad de brindar la información técnica que sirva de referencia para las futuras aplicaciones o que se utilice como una herramienta orientadora en la investigación educativa. Finalmente, poseer la verificación documental de la construcción del instrumento de evaluación (como lo es un Manual) también sirve para detectar las áreas de oportunidad en la mejora de los proyectos a corto plazo, al contener estándares para el diseño de pruebas.

La construcción de los manuales técnicos debe poseer evidencia que respalde los pasos metodológicos y los marcos conceptuales seguidos en cada periodo de aplicación de PLANEA EMS, por lo que el INEE recomienda que dichos documentos contengan los siguientes elementos:

- Antecedentes y los alcances del instrumento
- Diseño y construcción del instrumento
- Funciones desarrolladas por los cuerpos colegiados
- Administración del instrumento
- Características métricas de los reactivos
- Información sobre el proceso de puntuación
- Evidencias que sustentan la validez de las inferencias que se realizan, a partir de los resultados de la evaluación
- Indicadores de confiabilidad
- Error de muestreo al momento de declarar las inferencias que se pueden efectuar

- Evidencias posibles para sustentar la calidad métrica del instrumento
- Argumentos que permitan interpretar y emplear adecuadamente los resultados

Como se observa, cada uno de estos elementos ha sido desglosado y explicado a lo largo del presente Manual, por lo que se espera que sea un respaldo adecuado de la aplicación de PLANEA realizada en el 2017 y sirva como guía orientadora para posteriores procesos.

7.4. Revisiones posteriores a la aplicación definitiva de PLANEA EMS 2017

El inicio de un ciclo nuevo de aplicación de PLANEA EMS es el momento indicado para realizar una cuidadosa revisión de la prueba, que refleje las acciones tomadas para mantener un alto nivel académico del instrumento de evaluación. Por ello, una de las primeras acciones que se debe llevar a cabo es la revisión y, la actualización del instrumento de evaluación, principalmente en lo referente a la delimitación conceptual del objeto de medida y los contenidos a evaluar.

Así, para realizar esta revisión, es necesario que el Comité Académico de Determinación de Contenidos e Indicadores de Aprendizaje, cuerpo colegiado responsable de establecer el objeto de medida y describir los indicadores de aprendizaje a evaluar, se reúna antes de la nueva aplicación, para realizar una revisión de la Tabla de contenidos e indicadores de aprendizajes clave que se integró durante el diseño de PLANEA EMS 2017. La metodología seguida para el desarrollo de dicha labor, así como la documentación generada durante la sesión del Comité Académico constituyen un paso metodológico concerniente a la nueva aplicación y, por lo tanto, quedarán asentados en el Manual técnico que corresponda. Sin embargo, en este momento es necesario mencionar este paso para establecer una ruta de mejoramiento que sirva para la actualización del instrumento de evaluación.

Una vez que se ha revisado la delimitación conceptual, es preciso que dentro del plan de mejoramiento también se contemplen las especificaciones de reactivos, básicas para la construcción de los reactivos. Su revisión debe atender, principalmente, dos factores: los elementos de contenido y las características técnicas. En cuanto al primero, revisar que las especificaciones estén alineadas al marco conceptual y que representen el universo de medida (actualizado) de la prueba. Con respecto al segundo, que se observe su definición operacional (el conocimiento o habilidad a evaluar, el nivel de complejidad y la situación contextual), que se atienda al reactivo ejemplo, que se revisen tanto la respuesta correcta como los distractores y que se evalúe la pertinencia de la bibliografía de apoyo.

Una vez que se han integrado las revisiones anteriores, es imprescindible que se conforme un Consejo Académico Asesor de Educación Media Superior (CAAEMS), el cual tendrá la función de observar la revisión de la Tabla de contenidos que propuso el Comité Académico. Este CAAEMS será externo al INEE y estará conformado por autoridades educativas conocedoras de la EMS a partir de su experiencia como directores de planteles o subsistemas, secretarios académicos o personas que tengan algún puesto de gestión en los diferentes subsistemas de la EMS, de los bachilleratos de las universidades autónomas o de la iniciativa privada. La finalidad es que este Consejo realice una revisión experta del universo de medida de PLANEA EMS 2017 a partir de su Tabla de contenidos, para que pueda formular propuestas, observaciones y sugerencias de mejora al Comité Académico.
Por ello, resulta necesario que el Comité Académico vuelva a reunirse después del CAAEMS, para revisar e integrar, de ser el caso, las observaciones y sugerencias que éste propuso durante sus sesiones de trabajo.

Una constante en todos estos procesos será la implementación de estrategias para controlar el sesgo lingüístico y cultural. Así como en el diseño de PLANEA EMS 2017, en el que se procuró cuidar el sesgo desde la delimitación del universo de medida, de la misma manera, es recomendable que el mantenimiento y las revisiones subsecuentes del instrumento continúen por esta misma senda, para garantizar una prueba con equidad.

Finalmente, como apoyo a la documentación y resguardo de la información, los cambios importantes del instrumento de evaluación (como el ajuste al marco teórico, al objeto de medida o al tipo de inferencias que pueden realizarse a partir de los resultados), deben ser evidenciados en los informes y manuales técnicos de cada aplicación, pues éstos serán el respaldo histórico de PLANEA EMS 2017 y constituirán un apoyo para todos los usuarios de los resultados.

Conclusiones

La evaluación educativa es una herramienta para garantizar una educación de calidad con equidad para todos los ciudadanos. Bajo esta premisa, los instrumentos de evaluación contribuyen al mejoramiento de la educación obligatoria, pues con ellos se obtiene información necesaria para la toma de decisiones de política educativa. En este contexto, el Instituto Nacional para la Evaluación de la Educación (INEE), como autoridad en materia de evaluación, debe garantizar que los instrumentos que construye y diseña tengan solidez metodológica, técnica y conceptual, no sólo para obtener resultados confiables y válidos en términos del nivel de logro de los aprendizajes, sino también para que éstos puedan ser contextualizados desde las realidades de cada entidad federativa y adquieran una mayor relevancia para sus habitantes.

Uno de estos instrumentos es el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA), el cual fue puesto en práctica a partir del ciclo escolar 2014-2015 para evaluar el Sistema Educativo Nacional (SEN) e informar a las comunidades escolares del país sobre sus propios logros educativos. PLANEA es un instrumento de evaluación que retoma las experiencias anteriores en materia de evaluación (EXCALE y ENLACE), con la finalidad de mejorar la calidad de la educación mexicana. La aplicación de las pruebas se lleva a cabo al final del ciclo escolar en los grados terminales de todos los niveles de la educación obligatoria pues "se considera que los resultados de logro de aprendizaje al concluir un nivel educativo son un buen indicador de la eficacia del proceso educativo que ha transcurrido a lo largo de varios años de trabajo" (INEE, 2018: 20).

Así pues, con PLANEA se busca difundir resultados mediante los reportes institucionales y el informe nacional, elaborar bases de datos como apoyo a la investigación, publicar materiales informativos o informes temáticos que se centren en aspectos específicos de la evaluación educativa, con el objetivo de reflexionar y analizar los datos arrojados por la prueba.

Sin embargo, la evaluación a través de PLANEA no puede ser tomada como el único indicador del nivel de logro de los alumnos del país, ya que ésta refleja sólo un panorama a gran escala: "y no sustituye a las evaluaciones realizadas por cada docente en su práctica cotidiana" (INEE, 2016b: 47). Por esta misma razón, se requiere tener muy presente que los resultados de PLANEA significan un complemento y una retroalimentación, además de que: "son un referente más entre otros que sirven a escuelas y profesores para formarse una idea general del logro educativo a propósito de aquellos aspectos curriculares y habilidades específicas que fueron objeto de evaluación" (INEE, 2016b: 47).

En lo que respecta a la educación media superior (EMS), a partir de 2014 se incluyó la evaluación a gran escala de este nivel educativo, que anteriormente había sido evaluado mediante la prueba ENLACE. Las aplicaciones de PLANEA de los años 2014, 2015 y 2016 fueron coordinadas por la Secretaría de Educación Pública (SEP). A partir del 2017 el INEE asume la responsabilidad de diseñar, aplicar e informar sobre los resultados de PLANEA EMS. El objetivo es evaluar cada tres años a los estudiantes que concluyen la EMS a lo largo de la República mexicana, con la finalidad de ofrecer un margen de tiempo apropiado para que los resultados puedan mostrar cambios. De no hacerse así, sería difícil poder medir la evolución del SEN, dada su magnitud y la diversidad de los subsistemas que conforman la EMS.

Estas etapas en las aplicaciones también deben ir acompañadas de un apoyo para la verificación documental, en el cual se estipulen las fases del diseño de la prueba, sus actualizaciones y los cambios importantes tanto en sus referentes metodológicos como en los conceptuales. Por esta razón, un Manual técnico como el que se presenta ahora es una herramienta confiable pues sintetiza, describe y explica el diseño y las diversas fases por las que transita el instrumento de evaluación hasta la presentación de los resultados a nivel nacional. Es necesario que también se elabore un manual técnico después de cada aplicación de PLANEA SEN, ya que en éste deben incluirse los cambios que se realizaron tras cada actualización y aplicación. Lo anterior da cuenta del recorrido histórico de la prueba y permite demostrar su vigencia, al evidenciar que el instrumento de evaluación ha atendido los posibles cambios en el Marco Curricular Común (MCC) y en los planes y programas de estudio de la EMS.

Entonces, el Manual técnico de PLANEA EMS 2017 es el soporte documental del diseño de la prueba de 2017. Con este documento, el INEE contribuye con la transparencia de la información, evidencia los lineamientos y normativas que guiaron la construcción de la prueba, procura apoyar el mejoramiento de este instrumento de evaluación en sus subsecuentes aplicaciones, proporciona información para la toma de decisiones y la investigación educativa.

Como se ha descrito a lo largo del presente documento, PLANEA EMS 2017 abarcó el último grado de la EMS en escuelas de sostenimiento federal, estatal, autónomas y privadas. En este nivel educativo, se evaluaron aprendizajes clave de las áreas disciplinares de Lenguaje y Comunicación y Matemáticas, además de realizar la contextualización de las evaluaciones a través de cuestionarios para el alumno. Su diseño puso especial atención a la alineación de las pruebas con sus referentes curriculares, a la medición de los aspectos psicométricos, a la atención de la diversidad cultural y lingüística del país, a la metodología para la administración de la prueba, y a la amplia difusión de los usos de los resultados. Además, PLANEA EMS 2017 siguió los lineamientos establecidos por el INEE en sus documentos *Criterios técnicos para el desarrollo y uso instrumentos de evaluación* (2014) y *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación* (2017), los cuales fueron construidos con apego a la literatura especializada y a las recomendaciones internacionales.

El desarrollo de este Manual técnico, por lo tanto, también es una forma de vinculación más allá de los actores que tienen injerencia en la política educativa; también es un enlace con la población en general, en el cual se puede consultar los procesos que fueron seguidos y la forma en que se determinaron los resultados de los estudiantes del último año de la EMS. Además, se pueden comprobar los usos específicos de la prueba y descartar que éstos se utilicen con otros fines que no sean los de mejorar la educación. Debe recordarse que las evaluaciones del PLANEA no están diseñadas para: “evaluar el desempeño de los docentes, realizar rankings de escuelas, justificar decisiones de alto impacto o procesos de control administrativo sobre estudiantes, docentes o escuelas, tales como seleccionar alumnos, promover maestros o premiar escuelas” (INEE, 2018: 10). Finalmente, se debe recordar que se tienen proyectadas evaluaciones del SEN a nivel medio superior para los ciclos escolares 2020 y 2023, las cuales serán oportunamente divulgadas (tanto su diseño y aplicaciones como los resultados que arrojen).

Bibliografía

AERA-NCME-APA (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association / National Council on Measurement in Education.

Álvarez, P. (2007). *¿Qué es el modelo de Rasch?* Serie Estudios. 49. Santa Cruz Tenerife: Instituto Universitario de la Empresa / Universidad de la Laguna.

Burton, J., Moore, D. y Magliaro, S. (2004). Behaviorism and instructional technology. Jonassen, D. (Ed.). *Handbook of research on educational communications and technology.* Mahwah, NJ: Lawrence Erlbaum Associates, 3-36.

CENEVAL. Centro Nacional de Evaluación para la Educación Superior, A.C. (2013). *Manual técnico ENLACE Media Superior 2011-2012*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.enlace.sep.gob.mx/content/ms/docs/Manual_Tecnico_ENLACE%20MS_2011_2012.pdf

CENEVAL. (2016). *Manual técnico*. *Plan Nacional para la Evaluación de los Aprendizajes en el nivel medio superior. Planea MS 2015-2016*. (Versión preliminar en revisión). México: Autor.

CENEVAL. (s/f). *Estándares de calidad para instrumentos de evaluación educativa*. México: Autor.

Consejo Internacional de Monumentos y Sitios. (s/f). *Taxonomía de Bloom*. Recuperado el 3 de mayo de 2019 de: http://www.icomoscr.org/m/investigacion/%5BMETODOS%5DObjetivosTaxonomiaBloom.pdf

DOF. Diario Oficial de la Federación. (2008, 26 septiembre). *Acuerdo número 442 por el que se establece el Sistema Nacional de Bachillerato en un marco de diversidad*. México. Recuperado el 2 de mayo de 2019 de: http://www.sems.gob.mx/work/models/sems/Resource/10905/1/images/Acuerdo_numero_442_establece_SNB.pdf

DOF. (2012, 9 febrero). *Decreto por el que se declara reformado el párrafo primero; el inciso c) de la fracción II y la fracción V del artículo 3o., y la fracción I del artículo 31 de la Constitución Política de los Estados Unidos Mexicanos*. México. Recuperado el 2 de mayo de 2019 de: http://www.dof.gob.mx/nota_detalle.php?codigo=5233070&fecha=09/02/2012

DOF. (2017, 31 enero). *Estatuto Orgánico del Instituto Nacional para la Evaluación de la Educación*. México. Recuperado el 2 de mayo de 2019 de: http://www.dof.gob.mx/nota_detalle.php?codigo=5470412&fecha=31/01/2017

DOF. (2017, 15 septiembre). *Constitución Política de los Estados Unidos Mexicanos* (última reforma). México. Recuperado el 2 de mayo de 2019 de: http://www.diputados.gob.mx/LeyesBiblio/pdf/1_150917.pdf

DOF. (2018, 19 enero). *Ley General de Educación* (última reforma). México. Recuperado el 2 de mayo de 2019 de: https://www.sep.gob.mx/work/models/sep1/Resource/558c2c24-0b12-4676-ad90-8ab78086b184/ley_general_educacion.pdf

DOF. Diario Oficial de la Federación. (2018, 11 abril). *Criterios técnicos y de procedimiento para el análisis de los instrumentos de evaluación, el proceso de calificación y la definición de las listas de prelación de los concursos de oposición para el ingreso al Servicio Profesional Docente en educación básica y educación media superior para el ciclo escolar 2018-2019.* México: Autor.

DGLPD. Dirección General de Levantamiento y Procesamiento de Datos. (2017a). *Manual de procedimiento para la impresión y el empaquetado de materiales para el levantamiento de datos*. México: INEE.

DGLPD. (2017b). *Manual de procedimiento para el inicio formal del trabajo de campo para el levantamiento y procesamiento de datos en las entidades federativas.* México: INEE.

DGLPD. (2017c*). Reporte técnico. PLANEA EMS, ELSEN 2017*. México: INEE.

DGMTD. Dirección General de Medición y Tratamiento de Datos. (2017a). *Piloteo PLANEA 12, aplicado en 2016 (Resultados del análisis psicométrico de reactivos de logro de opción múltiple)*. México: INEE.

DGMTD. (2017b). *PLANEA 2017 Educación Media Superior. Diseño muestral*. México: INEE.

DGMTD. (2017c). *Análisis de reactivos PLANEA 2017 Educación Media Superior*. México: INEE.

DGMTD. (2018). *Metodología de escalamiento de PLANEA 2017 Educación Media Superior para la Evaluación de Logro en el Sistema Educativo Nacional (SEN).* México: INEE.

DGMTD. (2018b). *PLANEA EMS 2017. Reporte técnico de la construcción de índices y escalas de los datos de los cuestionarios de contexto.* México: INEE.

Félix, H*., et al.* (2009). Teoría de respuesta al ítem conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*. XVIII. 179-188.

Gipps, C. y Stobart, G. (2010). Fairness. Peterson, P., Baker, E., y MacGaw B. (eds.). *International Encyclopedia of Education*. 3ª ed. Oxford: Elsevier Academic Press, 56-60.

Gómez, J.; Hidalgo, M. y Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles del Psicólogo*. V. 31. (1). 75–84. Recuperado el 2 de mayo de 2019 de: http://www.redalyc.org/pdf/778/77812441008.pdf

Hernández, R., *et al*. (2014). *Metodología de la investigación*. 6ª ed. México: McGraw Hill.

Hopkins, K. (1998). *Educational and Psychological Measurement and Evaluation*. Boston: Allyn & Bacon.

INEE. Instituto Nacional para la Evaluación de la Educación (2005a). *Manual técnico de especificaciones de reactivos*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.inee.edu.mx/index.php/bases-de-datos/bases-de-datos-excale/marcos-de-referencia/95-publicaciones/de-pruebas-y-medicion-capitulos/471-manual-tecnico-especificaciones-de-reactivos

INEE. (2005b). *Manual técnico. Construcción de reactivos*. México: Autor. Recuperado el 2 de mayo de 2019 de: https://www.inee.edu.mx/index.php/bases-de-datos/bases-de-datos-excale/marcos-de-referencia/95-publicaciones/de-pruebas-y-medicion-capitulos/485-manual-tecnico-construccion-de-reactivos

INEE. (2006). *Manual técnico de validación de reactivos.* México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.inee.edu.mx/index.php/bases-de-datos/bases-de-datos-excale/marcos-de-referencia/95-publicaciones/de-pruebas-y-medicion-capitulos/484-manual-tecnico-validacion-de-reactivos

INEE. (2011). *La educación media superior en México. INFORME 2010-2011*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://publicaciones.inee.edu.mx/buscadorPub/P1/D/235/P1D235.pdf

INEE. (2013). *Ley del Instituto Nacional para la Evaluación de la Educación*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.inee.edu.mx/images/stories/2014/home/LeyINEE_web.pdf

INEE. (2014). *Criterios técnicos para el desarrollo y uso de instrumentos de evaluación educativa 2014-2015*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.inee.edu.mx/images/stories/2014/Normateca/2014/CRITERIOS_TECNICOS_PARA_EL_DESARROLLO_Y_USO_DE_INSTRUMENTOS_10_ABRIL_2014.pdf

INEE. (2015a). *Código de ética y conducta del Instituto Nacional para la Evaluación de la Educación*. México: Autor.

INEE. (2015b). *Plan Nacional para la Evaluación de los Aprendizajes. PLANEA. Documento rector.* México: Autor. Recuperado el 2 de mayo de 2019 de: http://planea.sep.gob.mx/content/general/docs/2015/PlaneaDocumentoRector.pdf

INEE. (2015c). *Política Nacional de Evaluación de la Educación. Documento rector*. México: Autor.

INEE. (2016a). *Panorama educativo de México. Indicadores del Sistema Educativo Nacional. Educación básica y media superior 2016.* México: Autor. Fecha de consulta en julio de 2018, de: http://publicaciones.inee.edu.mx/buscadorPub/P1/B/115/P1B115.pdf

INEE. (2016b). *PLANEA. Una nueva generación de pruebas.* México: Autor. Recuperado el 2 de mayo de 2019 de: http://publicaciones.inee.edu.mx/buscadorPub/P2/A/321/P2A321.pdf

INEE. (2016c). *Pautas editoriales para la construcción de reactivos*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.inee.edu.mx/images/stories/2014/Normateca/Pautas_Editoriales_para_la_construcci%C3%B3n_de_reactivos141216.pdf

INEE. (2017). *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://publicaciones.inee.edu.mx/buscadorPub/P1/E/104/P1E104.pdf

INEE. (2017b). *PLANEA. Resultados nacionales 2017*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://publicaciones.inee.edu.mx/buscadorPub/P2/A/328/P2A328.pdf

INEE. (2018). *Plan Nacional para la Evaluación de los Aprendizajes. PLANEA. Documento rector.* México: Autor. Recuperado el 2 de mayo de 2019 de: http://planea.sep.gob.mx/content/general/docs/2018/PlaneaDocumentoRector18.pdf

Madero, J. (2015, marzo). Configuración normativa de las leyes en el marco competencial de los órdenes jurídicos. *Congreso Redipal Virtual VIII*. México: Cámara de Diputados. Recuperado el 2 de mayo de 2019 de: http://www.diputados.gob.mx/sedia/sia/redipal/CRV-VIII-14-%2015.pdf

Martínez, F. (2008). Evaluación en gran escala y evaluación en aula. *Los usos de EXCALE para la mejora educativa*. México: INEE. Recuperado el 3 de mayo de 2019 de: http://www.inee.edu.mx/explorador/propuesta2008/Textos/EvaluacionGranEscala.pdf

Martínez, F. (coord.) (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación*. Cuaderno de investigación 40. México: INEE. Recuperado el 2 de mayo de 2019 de: http://publicaciones.inee.edu.mx/buscadorPub/P1/C/148/P1C148.pdf

UAM. Universidad Autónoma Metropolitana. (s/f). *La educación universitaria hoy.* México: UAM-Azcapotzalco. Recuperado el 3 de mayo de 2019 de: http://hadoc.azc.uam.mx/enfoques/conductismo.htm

UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2013). *Clasificación internacional normalizada de la educación*. Quebec: Autor.

SEP. Secretaría de Educación Pública. (2007). *Reforma integral de la educación media superior. La creación de un Sistema Nacional de Bachillerato en un marco de diversidad*. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.ofmx.com.mx/documentos/pdf/RIEMS_Creacion_Sistema_Nacional_de_Bachillerato.pdf

SEP. (2017a). *Planes de estudio de referencia del Marco Curricular Común de la Educación Media Superior*. México: Autor. Recuperado el 2 de mayo de 2019 de: https://www.gob.mx/cms/uploads/attachment/file/241519/planes-estudio-sems.pdf

SEP. (2017b). *Planes de estudio de referencia del Marco Curricular Común de la educación media superior*. México: Autor.

SEP. (s/f). Sistema Educativo Mexicano. México: Autor. Recuperado el 2 de mayo de 2019 de: http://www.mexterior.sep.gob.mx/sisedMEX.html

SEMS. Subsecretaría de Educación Media Superior (2013). *Antecedentes*. Recuperado el 2 de mayo de 2019 de: http://www.sems.gob.mx/en_mx/sems/antecedentes_dgb

SEMS. (2017). *Visión*. Recuperado el 2 de mayo de 2019 de: http://www.sems.gob.mx/es_mx/sems/mision_sems

Skinner, R. (1968). *A new format for ITV*. Educational Technology Research and Development, 16 (3), 287-293.

Torres, E., *et al*. (2002). *El concepto de competencia I. Una mirada interdisciplinar.* Bogotá: Sociedad Colombiana de Pedagogía.

Tristán, A. (s/f). *Rasch Anàlisis. Glosario Español*. Recuperado el 3 de mayo de 2019 de: https://www.rasch.org/rmt/glosario.htm

Triviño, Z., y Stiepovich, J. (2007). Indicadores de evaluación en la enseñanza-aprendizaje de enfermería. *Colombia Médica*, V. 38. Sup. 2. (4). Cali: Universidad del Valle, 89-97.

Wheeler, P., Haertel, G., y Scriven, M. (1992). *Teacher Evaluation Glossary*. Kalamazoo: CREATE Project / The Evaluation Center / Western Michigan University.

Glosario

Administración del instrumento: proceso en el que una o más personas responden el instrumento de evaluación.

Aprendizaje: adquisición de conocimientos, actitudes, valores o competencias por parte de una persona, y la generación de cambios en sus niveles de comprensión y comportamiento a través de la experiencia, la práctica, el estudio o la instrucción.

Aprendizaje clave: aprendizajes que favorecen la adquisición de otros aprendizajes en futuros escenarios de la vida y están comprendidos en los campos formativos y asignaturas, el desarrollo personal y social y la autonomía curricular. Los aprendizajes clave fortalecen la organización disciplinar del conocimiento y al mismo tiempo favorece su integración inter e intra asignaturas y de campos de conocimiento, a través de tres dominios organizadores: eje, componente y contenido central.

Área instrumental: áreas de aprendizaje que resultan esenciales para la continua adquisición de nuevos aprendizajes, son relevantes para el dominio del campo disciplinar y prevalecen en el tiempo a pesar de los cambios curriculares.

Atributo: caracteres cualitativos que en principio no pueden ser descritos numéricamente.

Bachillerato: se trata del único nivel que conforma a la educación media superior, comprende opciones que permiten siempre la formación para el ingreso a la educación superior y pueden, además, permitir la adquisición de habilidades laborales. Dichas opciones educativas son ampliamente diversas y son reguladas por el Sistema Nacional de Bachillerato y sujetas en términos curriculares al Marco Curricular Común.

Bachillerato general: nivel que está orientado a 1) consolidar los saberes adquiridos en etapas previas y 2) tener una función propedéutica que prepare a sus estudiantes para continuar sus estudios en el nivel superior.

Bachillerato tecnológico: nivel que oferta planes y programas orientados hacia a) capacitar al estudiante en el ejercicio de labores profesionales y b) tener una función propedéutica que prepare a sus estudiantes para continuar sus estudios en el nivel superior.

Banco de reactivos: repositorio donde se resguardan y clasifican los reactivos que integran los instrumentos de evaluación; en él se administran los datos de identificación del reactivo, sus características métricas, las formas en las que se incorporó y las fechas en las que se utilizó.

Calibración de reactivos: procedimientos para analizar las propiedades métricas de los ítems, verificando que cumplan con niveles aceptables según los propósitos de la prueba.

Cantidad de constructo: grado en el que está presente un atributo en un objeto de medición. Este atributo está asociado con algún constructo de interés. Depende de la forma en que sea definida la escala de medida para determinar si su valor es alto, medio o bajo dentro de una población de individuos u objetos de medición.

Campo disciplinar: áreas que comprenden contenidos y habilidades básicas y comunes para los estudiantes de educación media superior en cinco áreas de conocimiento: matemáticas, ciencias experimentales, comunicación, ciencias sociales y humanidades.

Competencias: movilización e integración de habilidades, conocimientos y actitudes en un contexto específico. El planteamiento curricular de la educación básica no parte de un enfoque explícito de aprendizaje por competencias; sin embargo, no por ello se deja de contemplar el desarrollo y práctica de conocimientos, habilidades, actitudes y valores. Precisamente, el Modelo Educativo considera poner más énfasis en asegurar que todos estos procedimientos y componentes logren consolidarse como aprendizajes más sólidos. En la educación media superior, esta estructura reordena y enriquece los planes y programas de estudio existentes y se adapta a sus objetivos; no busca remplazarlos, sino complementarlos y especificarlos. Define estándares compartidos que hacen más flexible y pertinente el currículo de la educación media superior. En este nivel son de tres tipos: genéricas, disciplinares y profesionales.

Competencias disciplinares: en la educación media superior comprenden las competencias básicas y extendidas.

Competencias disciplinares básicas: competencias comunes a todos los egresados de la educación media superior y representan la base común de la formación disciplinar en el marco del Sistema Nacional de Bachillerato.

Competencias disciplinares extendidas: competencias que son de mayor profundidad o amplitud que las competencias disciplinares básicas y dan especificidad a la oferta educativa proporcionada por los distintos subsistemas de la educación media superior.

Competencias genéricas: conocimientos, habilidades, actitudes y valores comunes a todos los egresados de la educación media superior, tienen aplicaciones diversas a lo largo de la vida, son transversales y transferibles en distintos espacios curriculares. Comprenden el autocuidado, la expresión, el pensamiento crítico, el aprendizaje autónomo, el trabajo colaborativo y la participación social.

Competencias profesionales: habilidades intelectuales y prácticas que están orientadas al ejercicio de un campo laboral específico, de acuerdo con las normas profesionales vigentes para cada opción. Se dividen en dos: las básicas, destinadas a establecer capacidades genéricas, y las extendidas, destinadas a desarrollar capacidades técnicas específicas.

Conductismo: enfoque educativo centrado en el profesor, el cual focaliza el trabajo docente conjuntamente con los planes de estudio, con una clara estructura jerárquica que no fomenta la actividad participativa del estudiante.

Confiabilidad: cualidad de las mediciones obtenidas con un instrumento, que se caracterizan por ser consistentes y estables cuando éste se aplica en distintas ocasiones.

Constructivismo: enfoque educativo que se centra en el estudiante y que propone un proceso en una bidirección permanente.

Constructo: objeto teórico de interés que se pretende medir en algún sujeto de estudio. Un constructo puede ser parte de un modelo teórico del conocimiento de una persona –tal como su comprensión de cierto conjunto de conceptos o su actitud hacia algo– o puede ser alguna variable psicológica tal como "necesidad de logro" o una variable de personalidad tal como un diagnóstico de bipolaridad. Puede ser del dominio del logro educativo o puede ser un constructo relacionado con la salud tal como "calidad de vida" o un constructo sociológico tal como "ruralidad" o grado de asimilación de los migrantes. También se puede relacionar con un grupo en lugar de con una persona, tal como trabajo de grupo o deportes de equipo o un ambiente de trabajo o puede ser un fenómeno biológico tal como la habilidad de un bosque de dispersarse en un nuevo ambiente. Puede incluso ser un objeto inanimado complejo como la proclividad de un volcán de hacer erupción o la erosión de muestras de pintura. Se utiliza como sinónimo de rasgo latente.

Correlación punto biserial: medida de consistencia que se utiliza en el análisis de reactivos; indica si hay una correlación entre el resultado de un reactivo con el resultado global del examen.

Criterio de evaluación: estándar o regla sobre la cual se puede establecer un juicio.

Cuestionario: conjunto de preguntas respecto de una o más variables que se van a medir.

Currículo: se concibe cada vez menos como una lista de contenidos y más como la suma y organización de parámetros que favorecen el desempeño de los alumnos y que dan lugar a una particular ecología del aprendizaje; es decir, a las relaciones simples y complejas que se producen entre los actores (alumnos, profesores, directivos, padres, autoridades, etc.) que participan del hecho educativo y a la interacción de éstos con el contexto del que forman parte. Entre los parámetros que dan forma al currículo destacan: ¿Para qué se aprende? ¿Cómo y con quién se aprende? y ¿Qué se aprende? Es decir, los fines, la pedagogía y los contenidos.

Desajuste: discrepancia existente entre los datos reales y los esperables según el cómputo global.

Desviación estándar: raíz cuadrada del promedio de las diferencias cuadráticas, obtenidas entre las medidas en lógitos y la media.

Dificultad de un reactivo: indica la proporción de personas que responden correctamente el reactivo de un examen.

Diseño: plan o estrategia que se desarrolla para obtener la información que se requiere en una investigación y responder al planteamiento.

Diseño matricial: diseño en el que se presenta a cada alumno un subconjunto del total de preguntas, eslabonando los bloques de reactivos, con la intención de evaluar un conjunto extenso de aprendizajes.

Distractor: opción de respuesta incorrecta del reactivo de opción múltiple, que probablemente será elegida por los sujetos con menor dominio en lo que se evalúa.

Educación: proceso mediante el cual las sociedades transmiten en forma intencionada el acervo de información, comprensión, conocimientos, actitudes, valores, habilidades, competencias y

comportamientos de una generación a otra. Comprende el uso de actividades de comunicación destinadas a producir aprendizaje.

Educación básica: tipo de educación que comprende los niveles de preescolar, primaria y secundaria en todas sus modalidades, incluyendo la educación indígena, la especial y la que se imparte en los centros de educación básica para adultos.

Educación formal: educación institucionalizada, intencionada y planificada por organizaciones públicas y organismos privados acreditados. En su conjunto, ésta constituye el sistema educativo formal del país. Por consiguiente, los programas de educación formal son reconocidos por las autoridades nacionales pertinentes o instancias equivalentes, por ejemplo, cualquier otra institución que colabore con las autoridades nacionales o subnacionales de educación. La educación formal comprende esencialmente la educación previa al ingreso al mercado laboral. Con frecuencia, la educación vocacional, la educación para necesidades especiales y parte de la educación de adultos se reconocen como parte integral del sistema nacional de educación formal.

Educación media superior: tipo de educación que comprende el nivel de bachillerato, los demás niveles equivalentes a éste, así como la educación profesional que no requiere bachillerato o sus equivalentes.

Educación obligatoria: se refiere a la suma de la educación básica y la educación media superior. El mandato establecido en el artículo 3º constitucional indica que es deber de todo mexicano atender a la educación básica y media superior. La educación básica comprende tres niveles educativos: preescolar con duración de tres años, primaria con duración de seis años y secundaria con duración de tres años. Abarca la formación de los tres a los 14 años de edad. La educación media superior se extiende por tres grados y comprende el rango de edad entre 14 y 17 años.

Educación por competencias: enfoque educativo centrado en el desempeño y que fomenta el desarrollo del pensamiento crítico, la resolución de problemas y el aprendizaje significativo, enfocado principalmente en la competencia dentro del campo laboral y, en general, en la vida socioafectiva del estudiante.

Eje temático: en la educación media superior organiza y articula los conceptos, habilidades, actitudes y valores de los campos disciplinares y es el referente para favorecer la transversalidad interdisciplinar.

Equiparación: método estadístico que se utiliza para ajustar las puntuaciones de las formas o versiones de un mismo instrumento, de manera tal que al sustentante le sea indistinto, en términos de la puntuación que se le asigne, responder una forma u otra.

Error de medida: diferencia entre el valor medido y el valor verdadero. Cuando la medida es más precisa, el error es más pequeño y viceversa.

Error estándar de medida: estimación de mediciones repetidas de una misma persona en un mismo instrumento que tienden a distribuirse alrededor de un puntaje verdadero. El puntaje verdadero siempre es desconocido porque ninguna medida puede ser una representación perfecta de un puntaje verdadero.

Escala: puntuaciones o medidas asignadas a objetos o sucesos a partir de un modelo de medición.

Escalamiento: proceso a través del cual se construye una escala.

Especificación de reactivos: descripción detallada de las tareas específicas susceptibles de medición, que deben realizar las personas que contestan el instrumento de evaluación. Deben estar alineadas al constructo definido en el marco conceptual.

Estándar: principio de valor o calidad en la conducción y uso de los procedimientos de evaluación y que son acordados por expertos en evaluación.

Evaluación: proceso sistemático mediante el cual se recopila y analiza información, cuantitativa o cualitativa, sobre un objeto, sujeto o evento, con el fin de emitir juicios de valor al comparar los resultados con un referente previamente establecido. La información resultante puede ser empleada como insumo para orientar la toma de decisiones.

Evaluación a gran escala: aplicación estandarizada de pruebas a grandes números de alumnos, para apreciar el nivel de aprendizaje que se alcanza en el sistema educativo de todo un país, región o distrito, ante la imposibilidad de agregar las evaluaciones que hacen los maestros, siempre ligadas al contexto en que trabaja cada uno.

Evaluación de logro: proceso sistemático y metódico mediante el cual se recopila información sobre el desempeño de un estudiante o grupo de estudiantes en relación con un conjunto de aprendizajes esperados y criterios o estándares de desempeño, con el fin de juzgar su mérito o valor para tomar decisiones específicas.

Evaluación educativa: emisión de un juicio basado en el análisis de evidencia sobre el estado de distintos aspectos de un proyecto educativo: procesos de enseñanza, contenidos, procesos de aprendizaje, procesos organizativos o el currículo, con el fin de tomar decisiones sobre el desarrollo posterior de los mismos.

Formas de un instrumento: dos o más versiones de un instrumento que se consideran equivalentes, pues se construyen con los mismos contenidos y especificaciones estadísticas.

Funcionamiento diferencial de los ítems: tendencia del reactivo a funcionar de manera diferente en diferentes subpoblaciones, a pesar de que los individuos que las componen obtengan puntuaciones similares en el reactivo.

Funcionamiento diferencial del instrumento: tendencia del instrumento a funcionar de manera diferente en diferentes subpoblaciones, a pesar de que los individuos que las componen obtengan puntuaciones similares en el instrumento. Las subpoblaciones son definidas por algo distinto a los aspectos relacionados con el constructo evaluado y suelen considerar aspectos de los individuos que las componen, tales como el sexo, la edad, el grupo étnico o el estatus socioeconómico.

Habilidad: nivel de desempeño exitoso con respecto al objeto de medida en la variable.

Indicador de aprendizaje: dato cualitativo o cuantitativo que, en conjunto con otros, muestra en qué condición se encuentra un objeto determinado y provee una base para establecer un juicio sobre el mismo.

Infit: estadístico de ajuste con información ponderada que se enfoca al comportamiento general de un ítem o de una persona. Se calcula con el promedio ponderado de las desviaciones (o diferencias) cuadráticas estandarizadas entre el desempeño observado y el esperado. El estadístico que se presenta en forma gráfica y tabular en el análisis de Rasch es la media cuadrática normalizada. Detecta desajustes en las desviaciones cerca de la zona de medición del ítem. El nombre "infit" es una abreviatura de "ajuste con información ponderada", por ello no se traduce al castellano. No debe traducirse como "ajuste interno". En algunos países latinoamericanos se denomina también como "ajuste próximo", que puede parecer más apropiado.

Instrumento: técnica que se utiliza para relacionar algo que observamos en el mundo real (a veces llamado manifiesto u observado) con algo que estamos midiendo que existe sólo como parte de una teoría (a veces llamado latente o inobservado). Este término es más amplio que el uso típico que se centra en la manifestación más concreta del instrumento: los reactivos o preguntas.

Instrumento de evaluación: herramienta de recolección de datos que suele tener distintos formatos, atendiendo a la naturaleza de la evaluación, por ejemplo, instrumentos de selección de respuesta, instrumentos de respuesta construida, cuestionarios, observaciones, portafolios, entre otros.

Instrumento de evaluación referido a un criterio: instrumento que permite comparar el desempeño de las personas evaluadas, con un estándar preestablecido.

IRT: teoría de respuesta a ítem o Teoría de respuesta al reactivo.

Jueceo: método en el cual se utiliza la opinión de jueces o expertos para determinar entre otras cosas: la pertinencia de la validez de los reactivos respecto a un dominio, el establecimiento de estándares o puntos de corte, la calificación de preguntas de respuesta no estructurada, de portafolios, etcétera.

Localización de alumnos o de reactivos: se asocia con los parámetros del Modelo de crédito parcial cuya estimación da información sobre la cantidad de constructo en un objeto de medición. También se usan los términos ubicación y posición como sinónimos de localización.

Mantenimiento: conjunto de procedimientos que tienen por objeto conservar actualizado el contenido de un instrumento de evaluación y vigilar su pertinencia, además de nutrir el banco de reactivos y tareas evaluativas con características cualitativas y cuantitativas óptimas.

Manual técnico: documento que contiene los elementos de fundamentación y planeación de un instrumento de evaluación, así como los resultados de los análisis de validez, confiabilidad y otros, necesarios para que se considere adecuado utilizarlo para los propósitos para los que fue diseñado.

Mapa: diagrama de barras (o histograma combinado) que muestra en dos partes de un mismo gráfico, a lo largo de la variable, la frecuencia y dispersión de los agentes (generalmente del lado derecho) contra la frecuencia y dispersión de los objetos (generalmente del lado izquierdo).

Marco Curricular Común: esquema de organización curricular para la educación media superior, que tiene como base las competencias genéricas, disciplinares (básicas y extendidas) y profesionales que deben desarrollar los estudiantes, y está orientado a dotar a la EMS de una identidad que responda a sus necesidades presentes y futuras.

Marco muestral: marco de referencia que permite identificar físicamente los elementos de la población, así como la posibilidad de enumerarlos y seleccionar las unidades muestrales.

Medición: asignación de números a categorías de observaciones. Las propiedades de los números se convierten en propiedades de la medición –nominal, ordinal, de intervalo o de razón–. El objetivo central de la medición es dar una manera congruente de resumir las respuestas que dan los alumnos para expresar su logro, actitudes o puntos de vista personales a través de instrumentos tales como escalas de actitud, pruebas de logro, cuestionarios, encuestas y escalas psicológicas.

Medida: ubicación de la variable latente (generalmente en unidades de lógitos). En el modelo de Rasch la medida de las personas se denomina "habilidad" y la medida de los ítems se denomina "dificultad".

MCP: modelo de crédito parcial

Modelo de Rasch: fórmula matemática para la relación entre la probabilidad de éxito (P) y la diferencia entre la habilidad del individuo (B) y la dificultad del ítem (D). $P= \exp (B-D)/(1+\exp(B-D))$ or $\log_e [P/(1-P)] = B – D$.

Muestra: subconjunto de la población de interés que refleja las variables medidas en una distribución semejante a la de la población.

Nivel de logro: criterio conceptual que delimita el marco interpretativo de las puntuaciones obtenidas en un instrumento de evaluación, y que refiere a lo que la persona evaluada es capaz de hacer en términos de conocimientos, destrezas o habilidades en el contexto del instrumento.

Objeto de medida: conjunto de características o atributos que se miden en el instrumento de evaluación.

Piloteo de reactivos: recolección preliminar de datos mediante la administración de un nuevo instrumento de evaluación para valorar su funcionamiento en una muestra de la población objetivo o en una población con características similares a las de la población objetivo, con el fin de hacer ajustes orientados a su mejora y a su posterior administración.

Planes y programas de estudio: programas donde se establecen los propósitos específicos de aprendizaje de las asignaturas u otras unidades de aprendizaje dentro de un plan de estudios, así como los criterios y procedimientos para evaluar y acreditar su cumplimiento. Podrán incluir sugerencias sobre métodos y actividades para alcanzar dichos propósitos. Dados los espacios de libertad y autonomía curricular contemplados en la educación básica, los programas pueden ser modificados en atención a las necesidades y contexto de los estudiantes, tras el debido análisis y discusión en los grupos colegiados.

Población objetivo: grupo de individuos sobre los cuales se desea que las inferencias elaboradas a partir de los resultados obtenidos con un instrumento de evaluación sean válidas.

Prueba: instrumento de evaluación que tiene como propósito medir el grado de dominio, conocimiento o aptitud para valorar el mérito de personas, instituciones, programas, sistemas, entre otros.

Punto de corte: en instrumentos de evaluación con referencia a un estándar de desempeño, es la puntuación mínima o el criterio a alcanzar o a superar para considerar que el nivel de desempeño de una persona cumple con lo esperado y distinguirlo de otro que no.

Puntaje: valores numéricos asignados a las respuestas cuando se suman para producir la calificación de un agente u objeto.

Puntuación: valor numérico obtenido durante el proceso de medición.

Rango: extensión total de los datos en la escala.

Rasgo latente: ver Constructo.

Reactivo o ítem: unidad básica de medida de un instrumento de evaluación que consiste en una pregunta o instrucción que requiere una respuesta del sujeto.

Respuesta: valor que indica el grado de éxito de un objeto o sujeto ante un agente.

Sesgo: error en la medición de un atributo debido a una variable no controlada, como las diferencias culturales o lingüísticas de las personas evaluadas.

Subsistema: para la educación media superior, cada una de las opciones que cuenten con sus propios planes de estudios, estándares de calidad, criterios de ingreso, egreso, objetivos y metas. Hasta hoy se cuentan más de treinta y pueden categorizarse de acuerdo con el tipo de institución que los administra, considérense, los centralizados del gobierno federal, los descentralizados del gobierno federal, los descentralizados de las entidades federativas, los estatales, los del gobierno de la Ciudad de México y los de la Universidades Autónomas.

Tasa: relación entre el número de casos de una categoría y el número total de observaciones.

Taxonomía: clasificación de los diferentes objetivos y habilidades educacionales que los docentes pueden proponer a los estudiantes. Tiene una estructura jerárquica que va del nivel más simple al más complejo.

Taxonomía de Bloom: esquema taxonómico de tres dominios: cognitivo, afectivo y psicomotor. La taxonomía de Bloom es jerárquica, es decir, asume que el aprendizaje a niveles superiores depende de la adquisición del conocimiento y habilidades de ciertos niveles inferiores. Al mismo tiempo, muestra una visión global del proceso educativo, promoviendo una forma de educación con un horizonte holístico.

Teoría de Respuesta al Ítem (TRI): relación entre el comportamiento de un sujeto frente a un ítem y el rasgo responsable de esta conducta (rasgo latente). El objetivo sustancial de la TRI es la construcción de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán igual probabilidad de dar la misma respuesta, independientemente de la población de pertenencia.

Universo de contenido: análisis formal del currículo que parte de un marco de referencia (planes y programas de estudio y, en el caso de la educación media superior, el Marco Curricular Común), para determinar la estructura formal de los posibles contenidos a evaluar.

Universo de medida: contenido que, con base en el diseño y objetivos de un instrumento de evaluación, es importante evaluar.

Usuarios: personas que toman decisiones con base en los resultados de la investigación; por ello, la presentación debe adaptarse a sus necesidades.

Validez: juicio valorativo integrador sobre el grado en que los fundamentos teóricos y las evidencias empíricas apoyan la interpretación de las puntuaciones de los instrumentos de evaluación.

Variable: idea o definición conceptual de lo que se desea medir. Una variable se define por medio de ítems o agentes de medida, usados para hacer explícitas sus manifestaciones o respuestas.

Variable latente: término estadístico empleado para denotar una característica o cualidad que se quiere medir en un objeto de medición pero que no es observable de manera directa.

Varianza irrelevante al constructo: efecto de variables ajenas al constructo que mide el instrumento de evaluación y que afecta sus resultados; por consiguiente, es información que compromete la validez de las inferencias que se realizan.

Varianza: la variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

ANEXOS

Anexo I. Perfiles

Perfiles profesionales y académicos de los integrantes en los Consejos y Comités que conforman la prueba PLANEA EMS 2017		
Diseño y estructura de la prueba		
Comité Académico	-Académico especialista en el campo disciplinar de Lenguaje y Comunicación o Matemáticas. -Investigador del nivel medio superior con acreditado sustento mediante publicaciones o trabajos referentes a este nivel escolar. -Docente en ejercicio, conocedor de los planes y programas de estudio del subsistema representado. -Conocedor del Marco Curricular Común desde la docencia, la investigación o la administración.	
Delimitación de la elaboración de los reactivos		
Comité de Especificaciones	-Experto en el Marco Curricular Común de las áreas disciplinares de Lenguaje y Comunicación o Matemáticas. -Especialista en didáctica con amplia experiencia docente en EMS o con experiencia en la formación de docentes. -Experto en el enfoque por competencias. -Autor de libros sobre temas del bachillerato. -Especialista del enfoque intercultural. -Conocedor del nivel medio superior desde la docencia, la investigación o la administración.	
Contrucción de los reactivos o ítems		
Comité de Elaboración de Reactivos	-Experto en el Marco Curricular Común de las áreas disciplinares de Lenguaje y Comunicación o Matemáticas. -Especialista en didáctica con amplia experiencia docente en EMS o con experiencia en la formación de docentes. -Experto en el enfoque por competencias. -Autor de libros sobre temas del bachillerato. -Especialista del enfoque intercultural. -Conocedor del nivel medio superior desde la docencia, la investigación o la administración.	
Revisión técnica, lingüística y cultural de los reactivos		
Comité de Validez y Sesgo	-Experto en el Marco Curricular Común de las áreas disciplinares de Lenguaje y Comunicación o Matemáticas. -Especialista en didáctica con amplia experiencia docente en EMS o con experiencia en la formación de docentes. -Experto en el enfoque por competencias. -Autor de libros sobre temas del bachillerato. -Especialista del enfoque intercultural. -Conocedor del nivel medio superior desde la docencia, la investigación o la administración.	
Determinación de niveles diferenciales de logro		
Comité de Niveles de Logro y Puntos de Corte	GRUPO A: -Especialista conocedor del currículo de la EMS, ya sea desde el MCC o los planes y programas de estudio de los diferentes subsistemas. -Investigador educativo del nivel medio superior. -Amplia experiencia en la elaboración de informes de resultados. GRUPO B: -Docente frente a grupo del nivel medio superior. -Autor de libros sobre temas del bachillerato. -Especialista en el campo disciplinar de Lenguaje y Comunicación o Matemáticas.	

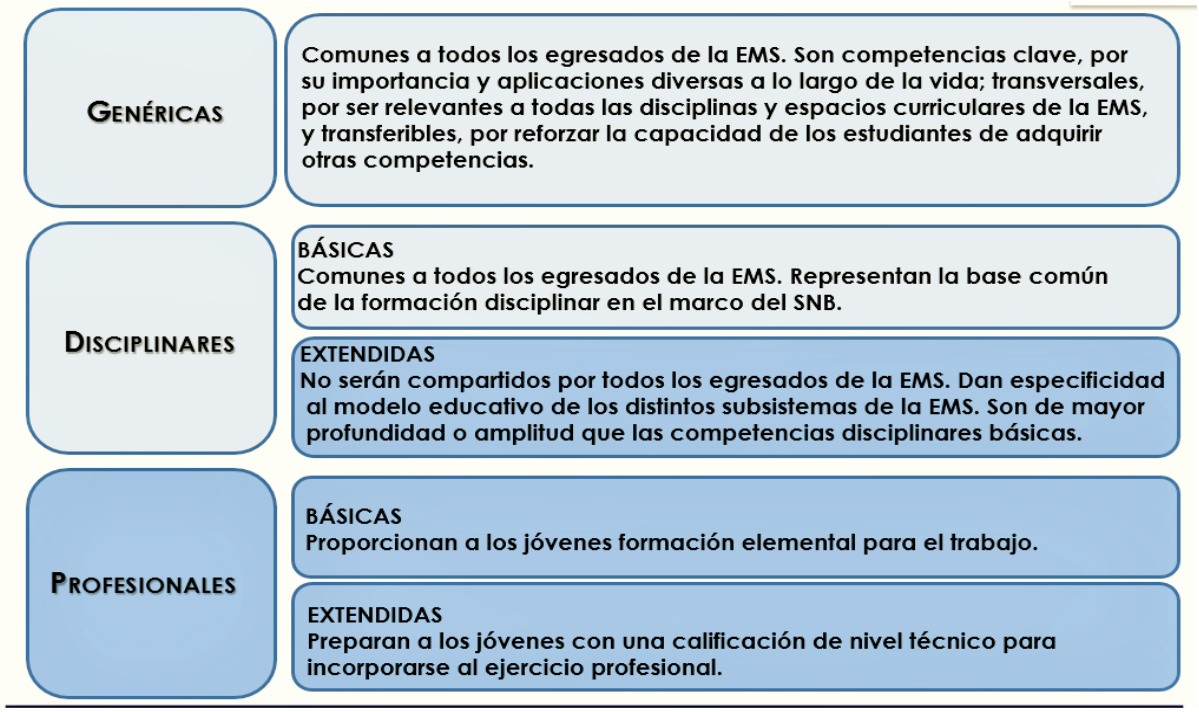
Revisión de la Tabla de contenidos

Consejo Académico Asesor de Educación Media Superior	-Conocedor del nivel medio superior desde la docencia, la investigación y la administración. -Formación como directivo, secretario académico o algún puesto de gestión en algún subsistema de EMS de la SEMS, de la iniciativa privada, o de las escuelas autónomas de reconocido prestigio.	

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS, INEE, México.

Anexo II. Materiales y formatos de trabajo del Comité Académico

Competencias genéricas, disciplinares y profesionales del MCC



Fuente: DOF. *Acuerdo número 442 por el que se establece el Sistema Nacional de Bachillerato en un marco de diversidad*.

Perfiles y competencias genéricas	
	Total de atributos
A. Se autodetermina y cuida de sí 1. Conoce y se valora a sí mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue. 2. Es sensible al arte y participa en la apreciación e interpretación de sus expresiones en distintos géneros. 3. Elige y practica estilos de vida saludables.	12
B. Se expresa y comunica 4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos mediante la utilización de medios, códigos y herramientas apropiados.	5
C. Piensa crítica y reflexivamente 5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos. 6. Sustenta una postura personal sobre temas de interés general, considerando otros puntos de vista de manera crítica y reflexiva.	10
D. Aprende de forma autónoma 7. Aprende por iniciativa e interés propio a lo largo de su vida.	3
E. Trabaja en forma colaborativa 8. Participa y colabora de manera efectiva en equipos diversos.	3
F. Participa con responsabilidad en la sociedad 9. Participa con una conciencia cívica y ética en la vida de su comunidad, región, México y el mundo. 10. Mantiene una actitud respetuosa hacia la interculturalidad y la diversidad de creencias, valores, ideas y prácticas sociales. 11. Contribuye al alcance de un equilibrio entre los intereses de corto y largo plazo con relación al ambiente.	15

Fuente: DOF. *Acuerdo número 442 por el que se establece el Sistema Nacional de Bachillerato en un marco de diversidad*.

Competencias disciplinares básicas del Marco Curricular Común*	
Campo disciplinar	**Disciplina**
Comunicación	Lectura y expresión oral y escrita, Taller de lectura y redacción, lengua adicional al español y Tecnologías de la Información y la Comunicación.
Matemáticas	Álgebra, Aritmética, Cálculo, Trigonometría y Estadística.
Ciencias experimentales	Física, Aritmética, Cálculo, Trigonometría y Estadística.
Humanidades y Ciencias sociales	Filosofía, Ética, Lógica, Estética, Derecho, Historia, Sociología, Política, Economía y Administración.

*Mencionadas en el artículo 7 del Acuerdo 656, por el que se adiciona el Acuerdo 444 establecen las competencias que constituyen el Marco Curricular Común del Sistema Nacional de Bachillerato y se adiciona el diverso 486 por el que se establecen las competencias disciplinares extendidas en el bachillerato general.

Competencias disciplinares de Matemáticas
1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
2. Formula y resuelve problemas matemáticos aplicando diferentes enfoques.
3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
4. Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
6. Cuantifica, representa y contrasta experimental o matemáticamente las magnitudes del espacio y las propiedades físicas de los objetos que lo rodean.
7. Elige un enfoque determinista o uno aleatorio para el estudio de un proceso o fenómeno, y argumenta su pertinencia.
8. Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Fuente: DOF. *Acuerdo número 442 por el que se establece el Sistema Nacional de Bachillerato en un marco de diversidad*.

Competencias disciplinares de Lenguaje y Comunicación
1. Identifica, ordena e interpreta las ideas, datos y conceptos explícitos e implícitos en un texto, considerando el contexto en el que se generó y en el que se recibe.
2. Evalúa un texto mediante la comparación de su contenido con el de otros, en función de sus conocimientos previos y nuevos.
3. Plantea supuestos sobre los fenómenos naturales y culturales de su entorno con base en la consulta de diversas fuentes.
4. Produce textos con base en el uso normativo de la lengua, considerando la intención y situación comunicativa.
5. Expresa ideas y conceptos en composiciones coherentes y creativas, con introducciones, desarrollo y conclusiones claras.
6. Argumenta un punto de vista en público de manera precisa, coherente y creativa.
7. Valora y describe el papel del arte, la literatura y los medios de comunicación en la recreación o la transformación de una cultura, teniendo en cuenta los propósitos comunicativos de distintos géneros.
8. Valora el pensamiento lógico en el proceso comunicativo en su vida cotidiana y académica.

9. Analiza y compara el origen, desarrollo y diversidad de los sistemas y medios de comunicación.

10. Identifica e interpreta la idea general y posible desarrollo de un mensaje oral o escrito en una segunda lengua, recurriendo a conocimientos previos, elementos no verbales y contexto cultural.

11. Se comunica en una lengua extranjera mediante un discurso lógico, oral o escrito, congruente con la situación comunicativa.

12. Utiliza las tecnologías de la información y comunicación para investigar, resolver problemas, producir materiales y transmitir información.

Fuente: DOF. Acuerdo número 442 por el que se establece el Sistema Nacional de Bachillerato en un marco de diversidad.

Matriz de doble entrada para la determinación de las competencias disciplinares que consolidan los atributos de las competencias genéricas

Competencias genéricas / **Competencias disciplinares**	**Perfil**					
	Competencia genérica					
Competencia disciplinar	Atributo	Atributo	Atributo	Atributo	Atributo	Atributo
Consolidación de los atributos						
Contenido temático						

Fuente: elaboración propia con datos de la Subdirección de Evaluaciones Nacionales de Resultados Educativos de EMS, INEE, México.

Anexo III. Materiales y formatos de trabajo del Comité de Especificaciones

Formato para la elaboración de las especificaciones de reactivos	
Coordinador académico:	
Redactores de la especificación:	
Revisores de la especificación:	
Fecha:	**Identificador único de la especificación:**

1. IDENTIFICACIÓN DEL CONTENIDO A EVALUAR

a. Campo disciplinar:
b. Competencia disciplinar:
c. Competencia genérica:
d. Atributo:
e. Contenido para la elaboración del reactivo (tema general del contenido a evaluar):
f. Indicador de aprendizaje:

2. DESCRIPCIÓN GENERAL DEL CONTENIDO A EVALUAR

- **a. Delimitación del contenido que se deberá cubrir:**
- **b. Interpretación del sentido del contenido:**
- **c. Importancia del contenido en función de la consolidación del atributo y la competencia disciplinar:**
- **d. Descripción de la función del contenido en el contexto de la disciplina:**

3. NIVEL DE DOMINIO

- **a. Nivel de dominio esperado:**
- **b. Conocimientos y habilidades requeridos para contestar correctamente el reactivo:**
- **c. Actividades cognoscitivas involucradas en la respuesta:**

4. PLANTILLA

- **Estructura de la base del reactivo:**
- **Estructura y descripción de la opción correcta:**
- **Estructura y descripción de las opciones consideradas incorrectas:**
- **Fuentes de consulta:**

5. REACTIVO MUESTRA

- **Estructura de la base del reactivo:**
- **Respuesta correcta:**
  A)
- **Distractores:**
  B)
  C)
  D)
- **Fuentes de consulta:**

6. FUENTES DE CONSULTA

(Según normas APA)

Revisión de especificaciones
Lista de cotejo

Fecha de la primera versión:	**Identificador único de la especificación:**
Fecha de la revisión:	**Número de revisión:**
Nombre del revisor de la especificación:	
Indicador de aprendizaje:	

Coordinador académico:

INSTRUCCIONES
Marque con un "X" las columnas SI o NO, según corresponda y describa las sugerencias para mejorar el rubro correspondiente.

Rubro	Criterios para la construcción de la especificación del reactivo	SÍ	NO	Observaciones
1. Datos de identificación	Todos los campos están completos.			

2. Descripción general del contenido a evaluar	Criterios para la construcción de la especificación del reactivo	SI	NO	Propuesta de corrección
a. Delimitación del contenido	Escribe el contenido temático.			
	Retoma la habilidad que establece la competencia disciplinar.			
	Describe el aspecto más relevante del aprendizaje observable.			
	Describe con argumentos teóricos los elementos de contenido.			
b. Interpretación del sentido del contenido	Describe la aplicación del tipo de aprendizaje (procedimental, conceptual, actitudinal o factual) en el contenido.			
	Acota las dificultades que presentan los estudiantes en este contenido.			
	Describe la relación directa entre el contenido y su aplicación en un contexto formativo integral.			

c. Importancia del contenido en función de la consolidación del atributo y la competencia disciplinar	Explicación de la correspondencia entre el atributo de la competencia genérica y las características del contenido.			
d. Descripción de la función del contenido en el contexto de la disciplina	Explica el impacto del contenido y del indicador dentro de la misma disciplina.			
	Explica la vinculación del indicador de aprendizaje a evaluar con otros aprendizajes o contenidos.			

3. Nivel de dominio	**Criterios para la construcción de la especificación del reactivo**	**SÍ**	**NO**	**Propuesta de corrección**
a. Nivel de dominio esperado	Indica los conocimientos, habilidades, actitudes y valores que acompañan a la ejecución del indicador de aprendizaje.			
	Argumenta los conocimientos, habilidades, actitudes y valores que acompañan a la ejecución del indicador de aprendizaje.			
b. Conocimientos y habilidades requeridos para contestar correctamente el reactivo	Enuncia las habilidades que se espera que el alumno tenga para responder el reactivo correctamente.			
	Enuncia las habilidades con el orden: verbo, tema y contextualización.			
c. Actividades cognoscitivas involucradas en la respuesta	Enuncia en función de la *Taxonomía de Bloom* las demandas cognoscitivas del indicador de aprendizaje.			

4. Plantilla	**Criterios para la construcción de la especificación del reactivo**	**SÍ**	**NO**	**Propuesta de corrección**
Estructura de la base del reactivo	Describe las características fundamentales que debe tener la base del reactivo para su elaboración.			
	El planteamiento de la base permite la elaboración de reactivos equivalentes.			
Estructura y descripción de la opción correcta	Describe las características necesarias para plantear la opción de respuesta correcta.			
Estructura y descripción de las opciones consideradas incorrectas	Los distractores que se proponen son plausibles de acuerdo con el contexto del problema.			

5. Reactivo muestra	**Criterios para la construcción de la especificación del reactivo**	**SÍ**	**NO**	**Propuesta de corrección**
Estructura del reactivo (base, opción correcta y distractores	El reactivo muestra, establece lo señalado en la plantilla de la especificación.			
	La base está conformada por una instrucción, aseveración o pregunta o una frase.			
	La redacción de la base del reactivo es clara precisa y concisa.			
	Utiliza una estructura sintáctica clara, coherente y adecuada.			
	Evita el uso de oraciones o enunciados condicionales.			
	Está libre de errores ortográficos.			

	Aplica las normas de elaboración de reactivos.			

6. Fuentes de consulta	**Criterios para la construcción de la especificación del reactivo**	**SÍ**	**NO**	**Propuesta de corrección**
En el reactivo muestra	Las referencias están redactadas siguiendo las normas del estilo APA.			
Para la especificación	Las referencias están redactadas siguiendo las normas del estilo APA.			

Anexo IV. Materiales y formatos de trabajo del Comité de Elaboración de Reactivos

Formato para la elaboración de reactivos

Nombre del redactor del reactivo (como aparece en el RFC):	**Campo disciplinar al que pertenece la prueba:**
Fecha de entrega solicitada (dd/mes/año):	
Número de entrega del reactivo: **Primera( ) Segunda( ) Tercera( ) Extraordinaria( )**	
Coordinador académico:	**Identificador único del reactivo:**

I. Identificación del contenido a evaluar

a. Competencia disciplinar: **b. Competencia genérica:** **c. Atributo:**	
Contenido temático:	**Indicador de aprendizaje como aparece en la especificación:**
Identificador único de la especificación:	

II. Reactivo completo

- **Estructura de la base del reactivo:**
- **Respuesta correcta:**
  A)
- **Distractores:**
  B)
  C)
  D)
- **Fuentes de consulta:**

III. Justificación de las opciones de respuesta

- **Respuesta correcta:**
  A)
- **Distractores:**
  B)
  C)
  D)

Formato para la revisión de reactivos
Lista de cotejo

I. Identificación del reactivo

Nombre del revisor del reactivo:	**Número de revisión del reactivo:** **Primera( ) Segunda( ) Extraordinaria( )**
Fecha de entrega solicitada:	**Fecha de entrega del revisor del reactivo:**
Identificador único del reactivo:	

II. Identificación del contenido a evaluar

Campo disciplinar:	**Competencia disciplinar – Competencia genérica:**
Contenido temático:	**Indicador de aprendizaje:**
Identificador único de la especificación:	

Evalúe los rubros de la columna de la izquierda que conforman el reactivo en su totalidad e indique si cumple o no con los criterios mencionados agregando los argumentos que justifiquen su opinión y añada una propuesta para corregir el error.

I. El reactivo como unidad	Sí	No	Argumentos	Propuesta de corrección
1. Se apega a la especificación.				
2. Se apega a la demanda cognitiva del indicador de aprendizaje.				
3. Es distinto al reactivo muestra de la especificación.				
4. Es una creación original del elaborador.				
5. Evalúa sólo un aprendizaje.				
6. Evalúa contenidos que no se respondan por sentido común ni por inferencia.				
7. Es independiente de otros reactivos, en el sentido que la información contenida en uno no sugiere la solución ni es requisito para contestar otro.				
8. Emplea un vocabulario claro y pertinente para la población objetivo.				
9. Está libre de expresiones idiomáticas locales que dificulten su comprensión.				
10. Está libre de errores de redacción y ortográficos.				
11. Está libre de texto o ilustraciones que puedan ser ofensivos o discriminatorios para algún grupo social.				
12. Omite estereotipos étnicos o de género y que no ponga en ventaja a un grupo social sobre otros.				
13. Considera únicamente la información necesaria y relevante para responder el reactivo.				
14. Utiliza opciones de respuesta distintas a la de otros reactivos.				
15. Permite evidenciar si el sustentante posee el conocimiento, la habilidad o la competencia que forma parte del objeto de medición del examen.				
16. Incluye la referencia documental correspondiente.				
17. La elaboración del reactivo está apoyada en fuentes de información oficiales y confiables.				
18. La extensión de las lecturas, cuando sean necesarias, considera como número máximo de palabras 600.				

II. La base del reactivo reactivo	Sí	No	Argumentos	Propuesta de corrección
1. Contiene un planteamiento que demande una tarea específica al individuo.				
2. Evalúa un contenido a través de una acción.				
3. Es un enunciado que plantea explícitamente un problema o una tarea.				
4. Especifica claramente la acción que se espera como respuesta.				
5. Está redactada como un enunciado interrogativo, como una afirmación o como una frase que se requiere completar.				
6. Está redactada en forma de enunciados, secuencias alfanuméricas, gráficas o imágenes en los que se omite uno o varios elementos, señalados con una línea.				

7.	Cuando se trate de lecturas para reactivos de completamiento, los espacios en blanco están todos del mismo tamaño (siete golpes de la barra espaciadora).				
8.	En caso de presentar un conjunto de elementos éstos pertenecen al mismo campo semántico.				
9.	Establece consignas directas y concisas.				
10.	Está redactada de forma que se entiende sin la necesidad de leer las opciones de respuesta.				
11.	Evita que la base del reactivo sea confusa debido al uso de lenguaje sofisticado.				
12.	Evita que la base del reactivo sea confusa debido al uso de una estructura gramatical o conceptual complejas.				
13.	Se abstiene del uso del condicional "SI" en la base.				
14.	Evita formular preguntas de manera ambigua.				
15.	Está libre de expresiones idiomáticas locales que dificulten su comprensión.				
16.	Usa verbos en imperativo al inicio de la instrucción, evitando mezclar el modo condicional y el modo interrogativo en una misma oración.				
17.	Contempla el uso de nombres propios representativos de México.				
18.	Utiliza nombres cortos sin diminutivos y sin apodos.				

	III. Las cuatro opciones de respuesta como unidad	**Sí**	**No**	**Argumentos**	**Propuesta de corrección**
1.	Tienen concordancia de género y número con la base del reactivo.				
2.	Tienen concordancia gramatical con la base.				
3.	Pertenecen al mismo tema o campo semántico.				
4.	Tienen el mismo nivel de generalidad o especificidad.				
5.	Son distintas entre sí; omiten el uso de sinónimos o respuestas equivalentes.				
6.	Cumplen con el requisito de no repetir frases o palabras entre las opciones.				
7.	Evita utilizar como opciones de respuesta: "todas las anteriores", "ninguna de las anteriores", "las dos de arriba", "A y C" o "no sé".				
8.	Omiten el uso de formas negativas o absolutas (no, nunca, siempre, completamente).				
9.	Evitan que una opción destaque, por su extensión, con respecto al resto.				
10.	Son menos extensas que la base del reactivo (salvo que el contenido o la especificación lo exijan).				

	IV. La opción de respuesta correcta	**Sí**	**No**	**Argumentos**	**Propuesta de corrección**
1.	Es la única correcta.				
2.	Evita dar pistas de ser la correcta.				
3.	Resuelve el planteamiento.				
4.	Es incontrovertible.				
5.	Está ubicada siempre en la primera opción.				
6.	Evita que la respuesta correcta se obtenga mediante procedimientos incorrectos.				
7.	Está argumentada de manera coherente considerando la base del reactivo.				

V. Los distractores	Sí	No	Argumentos	Propuesta de corrección
1. Presentan estímulos claros que no se presten a más de una interpretación.				
2. Evitan dar pistas que conduzcan a la respuesta correcta.				
3. Son elementos plausibles.				
4. Incluyen los errores más comunes de los sustentantes.				
5. Pertenecen al mismo campo semántico.				
6. Tienen relación con el planteamiento.				
7. Guardan la misma extensión que la respuesta correcta.				
8. Tienen congruencia gramatical con la pregunta.				
9. Están argumentados con suficiencia.				

Anexo V. Materiales y formatos de trabajo del Comité de Validez y Sesgo

Debido al carácter confidencial de los reactivos que fueron el principal insumo del Comité de Validez y Sesgo, se presenta solamente la lista de criterios que apoyaron la validación de los reactivos de PLANEA EMS 2017.

Guía para validación de reactivos
PLANEA EMS 2017

I. Contenido temático

1. **El reactivo es congruente con las competencias disciplinares y los programas de estudio del nivel que se evalúan**. Evalúa conocimientos y temas relacionados a los programas de estudio, competencias disciplinares y los libros de texto.
2. **Correspondencia con la especificación.** El reactivo evalúa lo señalado en la especificación y no otro aprendizaje.
3. **Demanda cognitiva de un aprendizaje específico.** Se resuelve sólo con los conocimientos y habilidades que se pretende evaluar.
4. **El reactivo tiene un nivel de dificultad adecuado.** Sin ser demasiado difícil o demasiado fácil presenta un nivel de dificultad apropiado para el grado escolar de los alumnos.

II. Inclusión

5. **Vocabulario.** Usa palabras cuyo significado es compartido entre los grupos sociales evaluados. Está escrito de modo que puede ser interpretado de igual forma por individuos de grupos sociales diversos y de diferentes partes de la República.
6. **Situación.** Contiene situaciones familiares a los alumnos de diversos grupos sociales.
7. **Aspectos relacionados con estereotipos de género.** Evita estereotipos de género, así como contextos asociados con los roles de género que faciliten la resolución de las preguntas sólo a hombres o sólo a mujeres.
8. **Culturales.** Evita la disonancia cultural y las generalizaciones respecto a estilo de vida, grupos etarios, subculturas, nivel de estudios, pertenencia regional, geográfica, o de migración que pudieran facilitar o dificultar la resolución del reactivo a algunos grupos sociales.
9. **Lingüísticas.** El lenguaje es claro, evita palabras, frases, expresiones de lenguaje figurado o retórico que dificultan la resolución de los reactivos, excepto cuando dichas expresiones son parte del contenido que se pretende medir.
10. **Socioeconómicas.** Evita mencionar características, estereotipos, atributos y significaciones que pudieran favorecer o dificultar la solución del reactivo a algunos grupos sociales.

III. Construcción de los reactivos

11. **Pertinencia conceptual global**. La base del reactivo, las opciones de respuesta, así como el texto auxiliar o gráfico, según el caso, son pertinentes en relación con los principios teóricos de la disciplina en que se fundamenta.
12. **Contexto apropiado al nivel educativo.** El contexto que se presenta en los reactivos es apropiado para los alumnos del grado escolar evaluado.
13. **Información necesaria.** La información que se presenta es la idónea en cantidad y en pertinencia.
14. **Tecnicismos**. En caso de utilizar tecnicismos, sólo emplea los necesarios para el contenido a medir.
15. **Pertinencia del texto auxiliar**. La lectura auxiliar al reactivo y el lenguaje empleado en éste son apropiados para los alumnos del grado escolar evaluado.
16. **Pertinencia del gráfico.** El diseño gráfico de las imágenes auxiliares a los reactivos es adecuado para los propósitos del reactivo. Las ilustraciones son claras y congruentes con la base del reactivo.
17. **Ortografía del texto auxiliar**. Cumple con las reglas ortográficas, como puntuación, acentuación y uso de mayúsculas.
18. **Redacción de la base del reactivo**. La redacción del planteamiento es sencilla y coherente. Se comprende la consigna.

19. **Ortografía de la base del reactivo.** Cumple con las reglas ortográficas, como puntuación, acentuación y uso de mayúsculas
20. **La repuesta señalada como correcta es la correcta.** La respuesta correcta señalada es efectivamente la correcta.
21. **Existe sólo una opción de respuesta correcta.** Sólo una opción cumple con todos los elementos para ser considerada una respuesta correcta.
22. **Redacción de las opciones de respuesta.** Hay consistencia y coherencia gramatical (sintaxis, género, número, persona, tiempo, modo, estructura y longitud) tanto entre la base y las opciones de respuesta del reactivo, como al interior de las opciones mismas.
23. **Se evitan pistas de la respuesta correcta.** Existe ausencia de pistas en la base del reactivo y en las opciones de respuesta.
24. **Otros.** Cualquier tipo de corrección que considere se deba realizar y que no esté contemplada en los puntos anteriores.

Anexo VI. Precisión del diseño muestral

La precisión en términos del valor de la razón del Error estándar del estimador de la media (EE) entre la Desviación estándar poblacional (SD): EE/SD debe ser menor o igual al 15% (DGMTD, 2017b: 23).

En este anexo se muestran las tasas de participación y el valor de la razón EE/SD. Los dominios sombreados en verde, tienen una tasa de participación de al menos 80% y el valor del cociente EE/SD< 15%; los dominios con tasas de participación menores a 80% están resaltadas en rojo; los dominios con el cociente EE/SD> 0.15 se encuentran resaltados en negritas (DGMTD, 2017b: 35).

Tabla por entidad, entidad-tipo de sostenimiento entidad-tipo de control administrativo

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
Aguascalientes	ENTIDAD		63	63	100	3830	3263	85	0	4.78	5.27
	Tipo de sostenimiento	Público	44	44	100	2850	2541	89	0	5.56	6
		Privado	19	19	100	980	722	74	1	8.45	8.89
	Tipo de control administrativo	Autónomo	4	4	100	304	285	94	0	7.77	6.64
		Estatal	21	21	100	1173	1062	91	0	5.66	6.93
		Federal	19	19	100	1373	1194	87	0	7.76	8.4
Baja California	ENTIDAD		74	74	100	4742	4057	86	0	8.54	7.57
	Tipo de sostenimiento	Público	53	53	100	3731	3204	86	0	10.76	9.78
		Privado	21	21	100	1011	853	84	0	12.45	10.58
	Tipo de control administrativo	Estatal	30	30	100	2067	1755	85	0	11.5	8.89
		Federal	**23**	**23**	**100**	**1664**	**1449**	**87**	**0**	**19.07**	**18.31**
Baja California Sur	ENTIDAD		55	55	100	3104	2652	85	0	3.7	4.88
	Tipo de sostenimiento	Público	39	39	100	2634	2236	85	0	3.65	5.07
		Privado	16	16	100	470	416	89	0	8.33	9.58
	Tipo de control administrativo	Estatal	25	25	100	1642	1465	89	0	3.89	4.61
		Federal	14	14	100	992	771	78	1	6.54	10.5
Campeche	ENTIDAD		57	57	100	2856	2472	87	0	4.9	5.24
	Tipo de sostenimiento	Público	40	40	100	2365	2094	89	0	5.28	5.59
		Privado	17	17	100	491	378	77	1	11.21	13.2
	Tipo de control administrativo	Autónomo	6	6	100	418	384	92	0	6.89	7.2
		Estatal	25	25	100	1312	1191	91	0	8.97	8.58

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
		Federal	9	9	100	635	519	82	0	9.34	12.41
Coahuila	ENTIDAD		77	76	99	4509	3817	85	0	8.56	8.71
	Tipo de sostenimiento	Público	49	48	98	3268	2829	87	0	7.15	5.74
		Privado	**28**	**28**	**100**	**1241**	**988**	**80**	**0**	**17.17**	**17.74**
	Tipo de control administrativo	Autónomo	13	13	100	922	735	80	0	6.02	6.53
		Estatal	15	15	100	918	866	94	0	9.58	9.17
		Federal	21	20	95	1428	1228	86	0	6.58	3.88
Colima	ENTIDAD		60	58	97	2945	2377	81	0	5.62	6.68
	Tipo de sostenimiento	Público	44	44	100	2374	2137	90	0	4.26	4.79
		Privado	16	14	88	571	240	42	1	18.61	23.56
	Tipo de control administrativo	Autónomo	21	21	100	1277	1164	91	0	6.64	7.08
		Estatal	13	13	100	531	474	89	0	6.72	7.65
		Federal	10	10	100	566	499	88	0	4.87	5.08
Chiapas	**ENTIDAD**		**69**	**62**	**90**	**3572**	**2918**	**82**	**0**	**18.8**	**15.12**
	Tipo de sostenimiento	**Público**	**49**	**42**	**86**	**2700**	**2243**	**83**	**0**	**19.47**	**15.72**
		Privado	20	20	100	872	675	77	1	15.84	18.62
	Tipo de control administrativo	**Estatal**	**34**	**29**	**85**	**1830**	**1568**	**86**	**0**	**20.19**	**16.21**
		Federal	15	13	87	870	675	78	1	20.55	19.37
Chihuahua	ENTIDAD		75	75	100	4563	3960	87	0	8.14	7.57
	Tipo de sostenimiento	Público	55	55	100	3653	3253	89	0	8.48	7.13
		Privado	20	20	100	910	707	78	1	22.44	23.08
	Tipo de control administrativo	Estatal	35	35	100	2264	1973	87	0	11.1	8.93
		Federal	20	20	100	1389	1280	92	0	7.81	10.11
Ciudad de México	ENTIDAD		42	40	95	2735	1871	68	1	13.85	15.06
	Tipo de sostenimiento	Público	23	23	100	1680	1229	73	1	21.47	22.26
		Privado	19	17	89	1055	642	61	1	21.4	22.53
	Tipo de control administrativo	Federal	23	23	100	1680	1229	73	1	21.47	22.26
Durango	ENTIDAD		68	68	100	3647	3188	87	0	5.89	5.17
	Tipo de sostenimiento	Público	48	48	100	2854	2482	87	0	6.48	5.61
		Privado	20	20	100	793	706	89	0	12.7	13.88
	Tipo de control administrativo	Autónomo	6	6	100	318	226	71	1	8.17	8.03
		Estatal	24	24	100	1206	1044	87	0	9.08	7.5
		Federal	18	18	100	1330	1212	91	0	9.69	9.93

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
Guanajuato	ENTIDAD		79	79	100	4855	4264	88	0	6.56	6.7
	Tipo de sostenimiento	Público	56	56	100	3734	3297	88	0	7.6	8.46
		Privado	23	23	100	1121	967	86	0	11.91	11.27
	Tipo de control administrativo	Autónomo	12	12	100	842	711	84	0	5.09	4.37
		Estatal	20	20	100	1140	1043	91	0	10.24	12.67
		Federal	24	24	100	1752	1543	88	0	5.9	5.69
Guerrero	ENTIDAD		92	92	100	5341	4034	76	1	4.94	5.53
	Tipo de sostenimiento	Público	76	76	100	4671	3558	76	1	5.23	5.87
		Privado	16	16	100	670	476	71	1	15.9	13.78
	Tipo de control administrativo	Autónomo	30	30	100	2065	1427	69	1	7.58	6.75
		Estatal	25	25	100	1172	1001	85	0	9.18	11.65
		Federal	21	21	100	1434	1130	79	1	11.97	12.26
Hidalgo	ENTIDAD		77	77	100	4769	4241	89	0	5.85	5.81
	Tipo de sostenimiento	Público	63	63	100	4047	3588	89	0	6.34	6.22
		Privado	**14**	**14**	**100**	**722**	**653**	**90**	**0**	**15.27**	**15.71**
	Tipo de control administrativo	Autónomo	18	18	100	1049	910	87	0	4.34	3.96
		Estatal	23	23	100	1322	1216	92	0	9.7	9.61
		Federal	22	22	100	1676	1462	87	0	8.16	7.85
Jalisco	ENTIDAD		120	119	99	7759	6546	84	0	5.96	6.48
	Tipo de sostenimiento	Público	98	98	100	6676	5752	86	0	5.87	6.4
		Privado	22	21	95	1083	794	73	1	20.1	21.4
	Tipo de control administrativo	Autónomo	55	55	100	3964	3450	87	0	7.32	7.5
		Estatal	**21**	**21**	**100**	**1163**	**956**	**82**	**0**	12.74	**18.02**
		Federal	22	22	100	1549	1346	87	0	5.08	5.24
México	ENTIDAD		104	103	99	6944	5640	81	0	4.66	5.6
	Tipo de sostenimiento	Público	81	80	99	5657	4681	83	0	5.58	6.54
		Privado	23	23	100	1287	959	75	1	11.23	9.46
	Tipo de control administrativo	Autónomo	20	19	95	1448	1164	80	0	3.29	3.15
		Estatal	36	36	100	2289	1980	87	0	6.31	7.99
		Federal	25	25	100	1920	1537	80	0	14	5.43
Michoacán	ENTIDAD		72	70	97	4087	3051	75	1	6.79	7.55
	Tipo de sostenimiento	Público	52	50	96	3000	2318	77	1	7.33	8.09
		Privado	20	20	100	1087	733	67	1	15.85	17.09
		Autónomo	14	14	100	882	533	60	1	5.19	5.4

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
	Tipo de control administrativo	Estatal	21	20	95	1018	875	86	0	10.62	11.42
		Federal	17	16	94	1100	910	83	0	10.27	12.98
Morelos	ENTIDAD		67	64	96	4292	3779	88	0	7.41	8.23
	Tipo de sostenimiento	Público	50	49	98	3404	3017	89	0	4.95	5.27
		Privado	**17**	**15**	**88**	**888**	**762**	**86**	**0**	**20.34**	**21.83**
	Tipo de control administrativo	Autónomo	13	13	100	916	790	86	0	3.82	3.38
		Estatal	21	20	95	1304	1184	91	0	6.87	4.25
		Federal	16	16	100	1184	1043	88	0	9.26	11.34
Nayarit	ENTIDAD		70	68	97	3560	2475	70	1	5.48	4.98
	Tipo de sostenimiento	Público	49	47	96	2610	2070	79	1	5.08	3.93
		Privado	21	21	100	950	405	43	1	15.57	15.73
	Tipo de control administrativo	Autónomo	20	19	95	1198	933	78	1	3.71	4.2
		Estatal	19	19	100	846	710	84	0	13.45	10.27
		Federal	10	9	90	566	427	75	1	10.09	8.29
Nuevo León	ENTIDAD		90	89	99	5977	4822	81	0	3.96	4.99
	Tipo de sostenimiento	Público	70	69	99	4836	4165	86	0	2.64	3.01
		Privado	20	20	100	1141	657	58	1	13	16.12
	Tipo de control administrativo	Autónomo	33	32	97	2410	2072	86	0	3.19	3.74
		Estatal	19	19	100	1211	995	82	0	4.79	8.15
		Federal	18	18	100	1215	1098	90	0	3.9	4.54
Oaxaca	ENTIDAD		74	73	99	3995	3259	82	0	5.64	6.42
	Tipo de sostenimiento	Público	59	59	100	3417	2769	81	0	6.16	7.15
		Privado	15	14	93	578	490	85	0	12.59	9.86
	Tipo de control administrativo	Autónomo	13	13	100	665	429	65	1	6.2	5.42
		Estatal	27	27	100	1376	1210	88	0	8.41	9.35
		Federal	19	19	100	1376	1130	82	0	7.36	11.36
Puebla	ENTIDAD		89	89	100	5340	4839	91	0	4.34	5.21
	Tipo de sostenimiento	Público	75	75	100	4661	4250	91	0	4.66	5.53
		Privado	14	14	100	679	589	87	0	13.43	**15.7**
	Tipo de control administrativo	Autónomo	23	23	100	1630	1473	90	0	5.91	6.77
		Estatal	35	35	100	1723	1599	93	0	5.86	7
		Federal	17	17	100	1308	1178	90	0	5.81	7.81
Querétaro	ENTIDAD		72	72	100	4561	3856	85	0	5.57	5.66
		Público	54	54	100	3604	3030	84	0	7.75	7.79

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
	Tipo de sostenimiento	Privado	18	18	100	957	826	86	0	3.71	4.76
	Tipo de control administrativo	Autónomo	10	10	100	663	568	86	0	5.23	5.52
		Estatal	33	33	100	2135	1798	84	0	11.23	11.44
		Federal	11	11	100	806	664	82	0	6.86	8.19
Quintana Roo	ENTIDAD		60	60	100	3472	3022	87	0	4.25	4
	Tipo de sostenimiento	Público	43	43	100	2689	2358	88	0	4.76	4.33
		Privado	17	17	100	783	664	85	0	11.22	10.51
	Tipo de control administrativo	Estatal	27	27	100	1617	1440	89	0	5.86	5.3
		Federal	16	16	100	1072	918	86	0	5.04	5.96
San Luis Potosí	ENTIDAD		51	51	100	3382	3137	93	0	11.55	13.93
	Tipo de sostenimiento	Público	41	41	100	2877	2672	93	0	6.5	6.49
		Privado	**10**	**10**	**100**	**505**	**465**	**92**	**0**	**25.29**	**30.25**
	Tipo de control administrativo	Autónomo	2	2	100	152	135	89	0	8.76	8.32
		Estatal	26	26	100	1771	1646	93	0	8.22	8.7
		Federal	13	13	100	954	891	93	0	10.09	7.48
Sinaloa	ENTIDAD		104	103	99	6078	5162	85	0	3.5	3.44
	Tipo de sostenimiento	Público	85	85	100	5329	4488	84	0	3.5	3.58
		Privado	**19**	**18**	**95**	**749**	**674**	**90**	**0**	**16.95**	13.95
	Tipo de control administrativo	Autónomo	42	42	100	2660	2034	76	1	5.77	5.06
		Estatal	23	23	100	1243	1109	89	0	6.17	7.82
		Federal	20	20	100	1426	1345	94	0	5.36	6.45
Sonora	ENTIDAD		65	65	100	4092	3571	87	0	5.39	5.49
	Tipo de sostenimiento	Público	45	45	100	3073	2691	88	0	5.93	5.95
		Privado	20	20	100	1019	880	86	0	13.45	14.12
	Tipo de control administrativo	Estatal	26	26	100	1708	1558	91	0	7.09	6.69
		Federal	19	19	100	1365	1133	83	0	11.35	12.45
Tabasco	ENTIDAD		77	76	99	4666	4041	87	0	4.16	4.03
	Tipo de sostenimiento	Público	60	60	100	3942	3402	86	0	4.54	4.42
		Privado	17	16	94	724	639	88	0	14.04	12.08
	Tipo de control administrativo	Estatal	45	45	100	2879	2553	89	0	5.3	5.29
		Federal	15	15	100	1063	849	80	0	5.97	4.9
Tamaulipas	ENTIDAD		72	72	100	4842	4197	87	0	6.33	6.39
		Público	56	56	100	3890	3454	89	0	5.88	5.47

Entidad	Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
	Tipo de sostenimiento	Privado	16	16	100	952	743	78	1	16.34	17.33
	Tipo de control administrativo	Autónomo	5	5	100	342	251	73	1	6.26	6.74
		Estatal	21	21	100	1364	1190	87	0	11.53	10.68
		Federal	30	30	100	2184	2013	92	0	6.38	6.07
Tlaxcala	ENTIDAD		59	59	100	3363	2952	88	0	5.08	4.23
	Tipo de sostenimiento	Público	42	42	100	2797	2456	88	0	5.66	4.6
		Privado	17	17	100	566	496	88	0	12.21	11.84
	Tipo de control administrativo	Estatal	27	27	100	1721	1535	89	0	7.74	6.23
		Federal	15	15	100	1076	921	86	0	4.49	5.48
Veracruz	ENTIDAD		80	80	100	4231	3883	92	0	7.88	8.02
	Tipo de sostenimiento	Público	60	60	100	3400	3150	93	0	8.96	9.08
		Privado	20	20	100	831	733	88	0	13.89	13.73
	Tipo de control administrativo	Estatal	34	34	100	1549	1448	93	0	11.85	12.08
		Federal	26	26	100	1851	1702	92	0	6.69	8.33
Yucatán	ENTIDAD		63	63	100	3745	3174	85	0	5.54	6.17
	Tipo de sostenimiento	Público	46	46	100	2840	2377	84	0	6.11	6.98
		Privado	17	17	100	905	797	88	0	8	9.03
	Tipo de control administrativo	Autónomo	4	4	100	304	270	89	0	6.81	7.32
		Estatal	27	27	100	1585	1347	85	0	8.68	10.1
		Federal	15	15	100	951	760	80	0	6.49	6.89
Zacatecas	ENTIDAD		67	67	100	3662	3180	87	0	4.16	4.98
	Tipo de sostenimiento	Público	50	50	100	3083	2661	86	0	4.44	5.35
		Privado	17	17	100	579	519	90	0	11.47	8.93
	Tipo de control administrativo	Autónomo	16	16	100	937	746	80	0	4.4	5.18
		Estatal	22	22	100	1272	1136	89	0	6.64	8.35
		Federal	12	12	100	874	779	89	0	8.07	6.9

Fuente: Dirección General de Medición y Tratamiento de Datos.

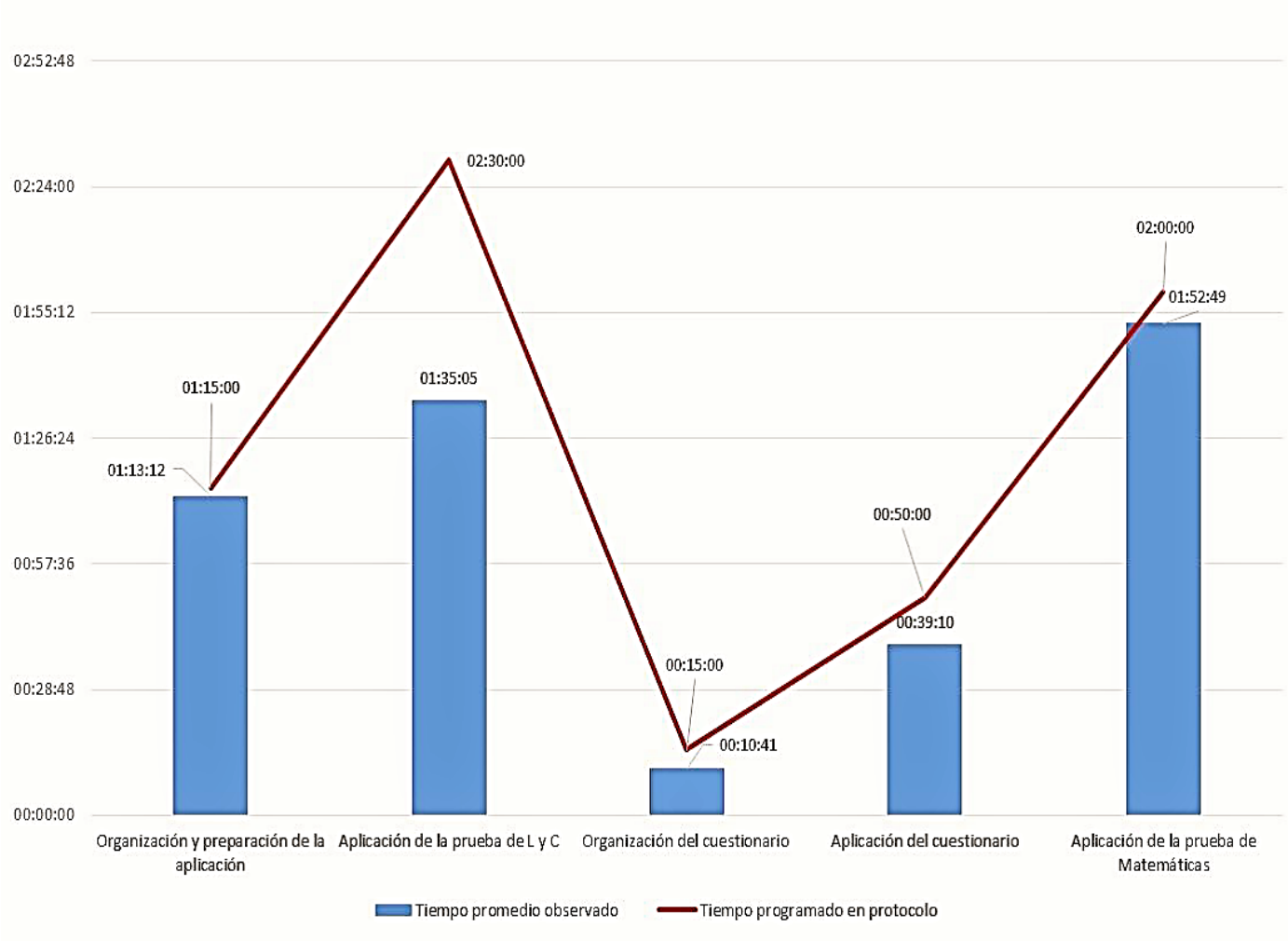
Tabla por tipo de sostenimiento, tipo de control administrativo y tipo de servicio

Dominios de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final LyC	Precisión final Mat
Tipo de sostenimiento	Público	1755	1740	99	112346	96452	86	0	3.2	2.5
	Privado	589	579	98	27170	21248	78	1	3.6	3.9
Tipo de control administrativo	Autónomo	380	377	99	25366	20690	82	0	2.4	2.5
	Estatal	800	793	99	46351	40927	88	0	4.2	3.5
	Federal	575	570	99	40629	34835	86	0	6.1	4.5
Tipo de servicio	Bachillerato estatal	92	91	99	5020	4326	86	0	5.2	6.4
	Bachillerato autónomo	380	377	99	25366	20690	82	0	2.4	2.5
	CECYTE	**140**	**137**	**98**	**9263**	**8395**	**91**	**0**	**18.7**	**15.3**
	Colegio de Bachilleres	277	277	100	19014	16684	88	0	5.5	2.6
	CONALEP	124	124	100	8560	7339	86	0	3.6	3.2
	DGETA	89	88	99	5670	4902	86	0	5.6	5.2
	DGETI	368	365	99	27189	23624	87	0	2.3	2.7
	Telebachillerato	36	34	94	1137	1016	89	0	8.1	9.1
	Telebachillerato comunitario	52	52	100	862	793	92	0	7.5	7.1

Nota: Los dominios sombreados en verde, tienen una tasa de participación de al menos del 80% y el valor del cociente EE/SD<= 15%; los dominios con tasas de participación menores al 80% están resaltadas en rojo; los dominios con el cociente EE/DE > 0.15 se encuentran resaltados en negritas.
Fuente: Dirección General de Medición y Tratamiento de Datos.

Anexo VII. Promedio de las actividades de la aplicación nacional

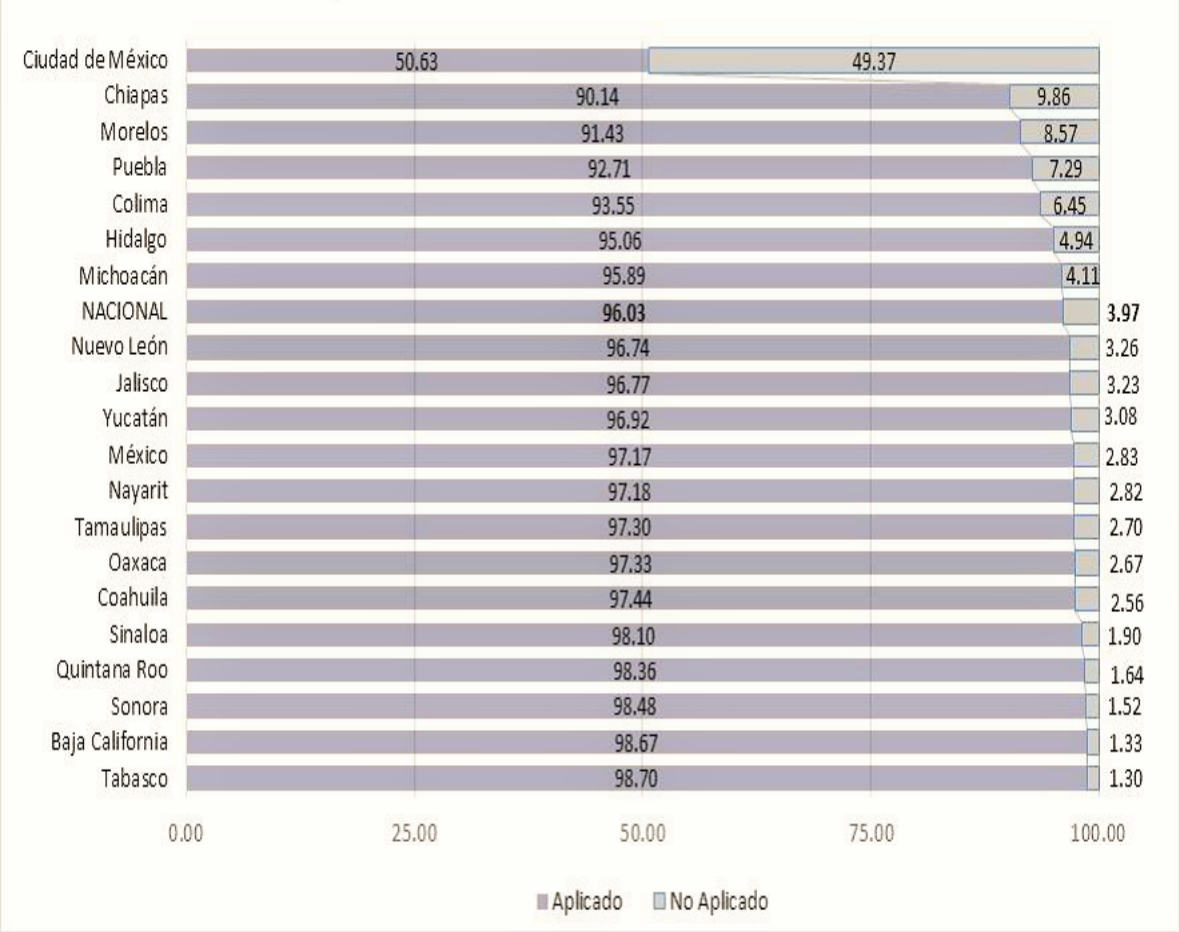
Duración promedio de las actividades de la aplicación con respecto al protocolo



Fuente: *Reporte Técnico PLANEA EMS, ELSEN 2017*, de la DGLPD.

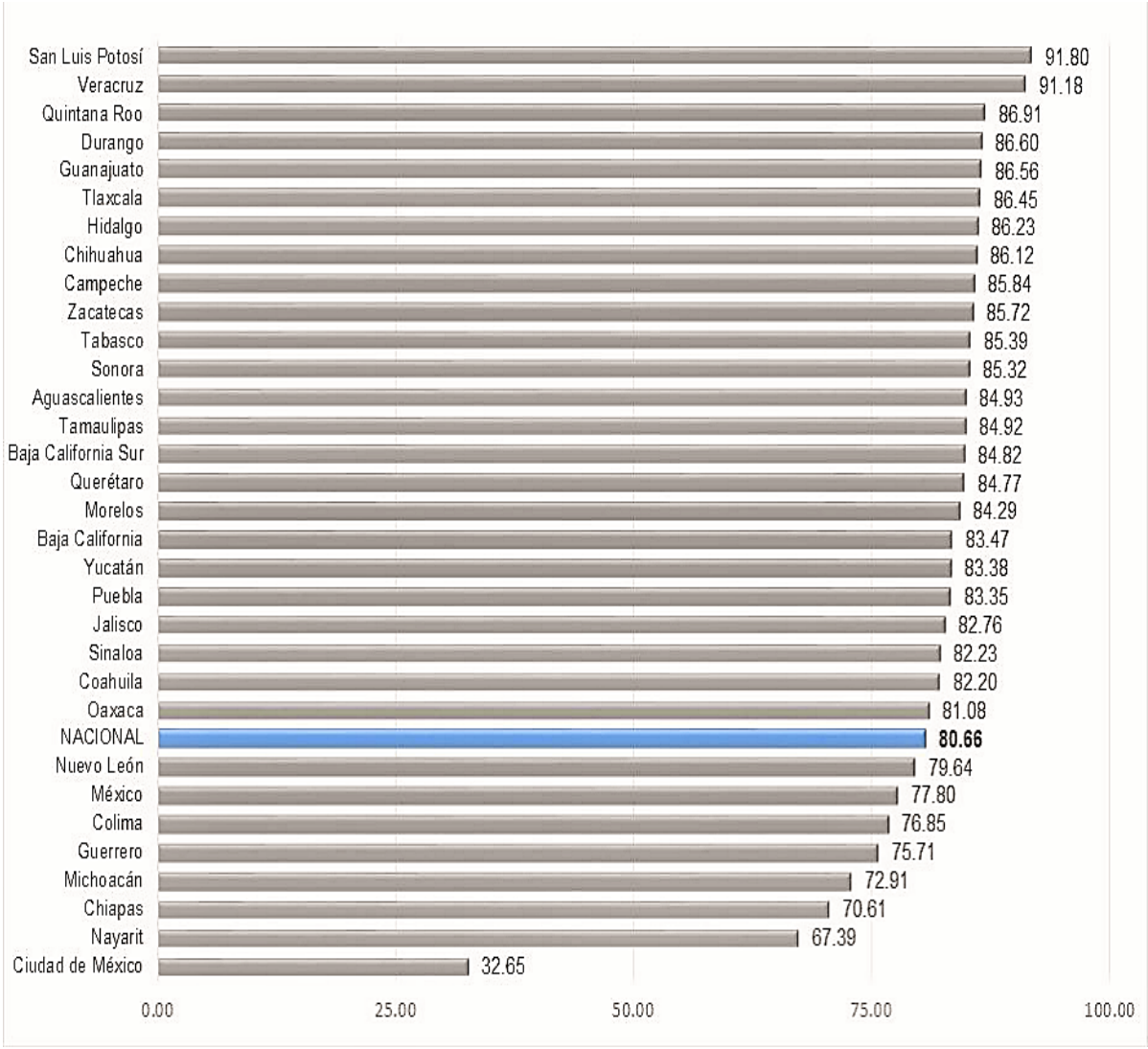
Anexo VIII. Porcentaje de aplicación por planteles-turno PLANEA EMS 2017

Tasa nacional de aplicación



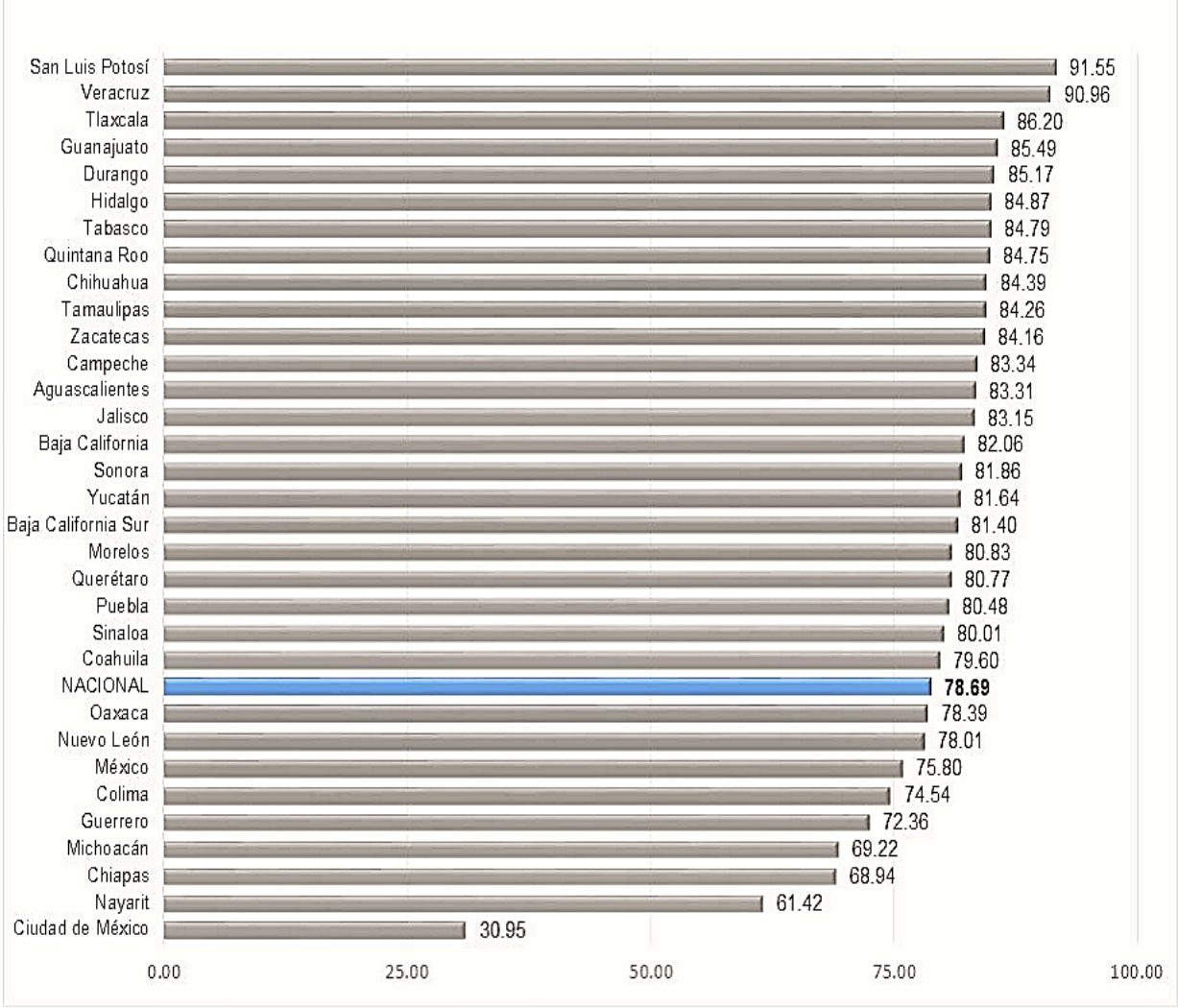
Fuente: *Reporte Técnico PLANEA EMS, ELSEN 2017*, de la DGLPD.

Anexo IX. Porcentaje de respuesta en Lenguaje y Comunicación



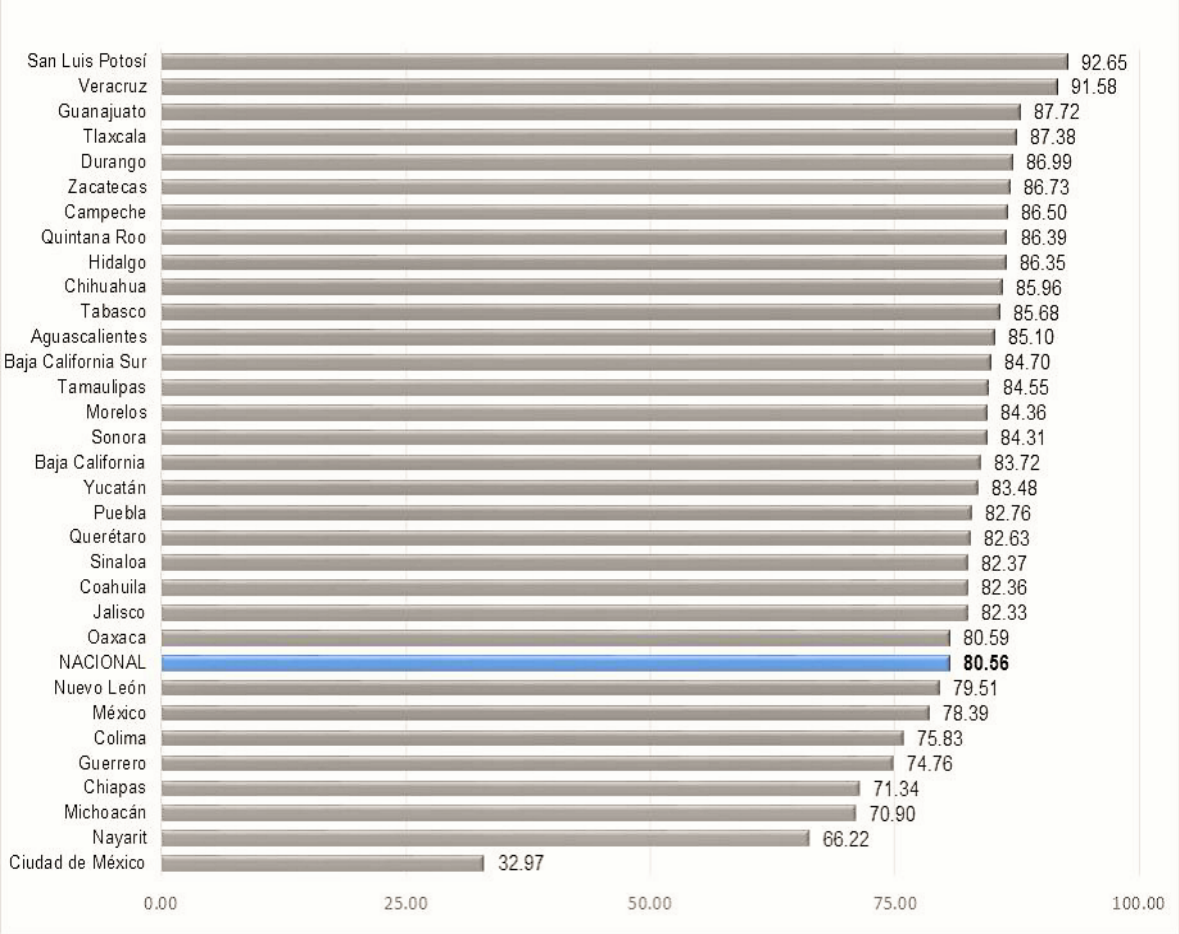
Fuente: *Reporte Técnico PLANEA EMS, ELSEN 2017*, de la DGLPD.

Anexo X. Porcentaje de respuesta en Matemáticas



Fuente: *Reporte Técnico PLANEA EMS, ELSEN 2017*, de la DGLPD.

Anexo XI. Porcentaje de respuesta del Cuestionario para el alumno



Fuente: *Reporte Técnico PLANEA EMS, ELSEN 2017*, de la DGLPD.

Anexo XII. Escalamiento de logro PLANEA EMS 2017 (tratamiento de los datos faltantes en los reactivos de logro)

Los reactivos que se utilizaron en PLANEA EMS 2017 para medir el logro de los estudiantes en las asignaturas de Lenguaje y Comunicación y Matemáticas fueron de opción múltiple. Para cada una de estas opciones de respuesta se asociaron los números del 1 al 4.

En las evaluaciones de gran escala pueden ocurrir dos situaciones:

1. Los sustentantes respondieron una y sólo una de las posibles opciones de respuesta y se les asignó un puntaje tratándolos como reactivos dicotómicos, es decir, 1 si respondió correctamente y 0 cuando respondió incorrectamente.
2. Los sustentantes dejan de responder por alguna razón o responden más de una de las opciones de respuesta.

La segunda situación nunca es deseada, pues genera respuestas faltantes, sin embargo, es inevitable tener que lidiar con ella. Para PLANEA SEN, éstos y otros tipos de respuestas faltantes se consideraron como valores perdidos que recibieron un tratamiento distinto en los diferentes momentos del proceso de construcción de las escalas.

Como primer paso, se presentan las distintas categorías de respuestas faltantes consideradas en PLANEA SEN para los reactivos correspondientes al logro:

Respuesta omitida: ocurrió cuando un examinado tuvo el tiempo y la oportunidad de responder a un reactivo, sin embargo, no lo hizo. Se le asignó el código ‘9’.

Respuesta múltiple: se consideró cuando un examinado respondió más de una de las opciones de respuesta. Se le asignó el código ‘8’.

No administrada: las pruebas de logro se aplicaron mediante un esquema de rotación de bloques de reactivos siguiendo un esquema matricial, de tal forma que a cada examinado se le asignó un subconjunto a responder de toda la prueba. A las respuestas de los reactivos que no se les asignó para responder, se les consideró de esta clase y se les asignó el código ‘7’.

No alcanzada: bajo las mismas condiciones de aplicación y tiempo, algunos examinados no alcanzaron a responder los reactivos del final de la sesión. A estas respuestas faltantes se les asigno el código ‘5’.
Una vez identificados los distintos tipos de datos faltantes, se les asignó un puntaje diferente en cada una de las etapas de la construcción de las escalas. A continuación, se describe la asignación.

Puntajes para las respuestas faltantes:
Para asignar los puntajes de las respuestas faltantes se siguió la estrategia sugerida por H. Ludlow y O'Leary (1999) en la que el tratamiento de las respuestas faltantes se hace diferenciado en un proceso de dos etapas.

Esta estrategia es la que usan en la actualidad las pruebas TIMSS en su proceso de escalamiento (Adams, R.J.; Wu, M.L.; Macaskill, G., 1997).

En la primera etapa, en la que se hizo la estimación de los parámetros de dificultad de los reactivos bajo el modelo de Rasch (calibración), las respuestas no alcanzadas se trataron como no administradas, por lo que no se les asignó un puntaje.

La razón que describen H. Ludlow y O'Leary (1999) es la siguiente: si en este punto del proceso de estimación suponemos que las respuestas no alcanzadas fueran tratadas como incorrectas, ocurriría una sobreestimación de la dificultad de los reactivos, ya que los estudiantes podrían contestar algunas de estas preguntas correctamente, teniendo el tiempo y la oportunidad de responder a ellas.

En la segunda etapa, se consideraron fijos y conocidos los parámetros de dificultad de los reactivos y se hizo la estimación de las habilidades de los examinados. En esta etapa, las respuestas no alcanzadas se consideraron como respuestas incorrectas, por lo que se les asignó un puntaje de '0'. H. Ludlow y O'Leary (1999) dicen que, si suponemos que las respuestas no alcanzadas son tratadas como no administradas, en la fase de estimación de las habilidades, los estudiantes podrían adoptar la estrategia de ir respondiendo desde el inicio únicamente los reactivos en los que está seguro de conocer la respuesta correcta y detenerse en el momento en el que ya no se sienta seguro y dejar el resto de la prueba en blanco. Como las respuestas no alcanzadas no serían tomadas en cuenta para la estimación de la habilidad, una gran parte de ellos podría obtener muy buenos resultados en la prueba. Por lo que, en ambas etapas, las respuestas omitidas y múltiples se consideraron como respuestas incorrectas, por lo que se les asignó un puntaje de '0'.

El mecanismo de administración de los reactivos, el cual se hizo mediante la rotación de los bloques bajo un diseño matricial, hace posible ignorar las respuestas no administradas, por lo que no se les asignó un puntaje, se les consideró '*missing by design*' (Boomsma, A., van Duijn, y Snijders, A. 2001) (DGMTD, 2018: 17-18).

Anexo XIII. Materiales y formatos de trabajo del Comité de Niveles de Logro y Puntos de Corte

Descriptores genéricos de los niveles de logro			
Nivel I	**Nivel II**	**Nivel III**	**Nivel IV**
Los estudiantes que se ubican en este nivel obtienen puntuaciones que representan un **logro insuficiente de los aprendizajes** clave, lo que refleja **carencias fundamentales** para seguir aprendiendo.	Los estudiantes que se ubican en este nivel tienen un **logro apenas indispensable de los aprendizajes** clave.	Los estudiantes que se ubican en este nivel tienen un **logro satisfactorio de los aprendizajes** clave.	Los estudiantes que se ubican en este nivel tienen un **logro sobresaliente de los aprendizajes** clave.

Formato de niveles de logro Lenguaje y Comunicación PLANEA EMS 2017		
Nivel I	Manejo y construcción de la información	
	Texto argumentativo	
	Texto expositivo	
	Texto literario	

Nivel II	Manejo y construcción de la información	
	Texto argumentativo	
	Texto expositivo	
	Texto literario	

Nivel III	Manejo y construcción de la información	
	Texto argumentativo	
	Texto expositivo	
	Texto literario	

Nivel IV	Manejo y construcción de la información	
	Texto argumentativo	
	Texto expositivo	
	Texto literario	

Formato de niveles de logro Matemáticas **PLANEA EMS 2017**		
Nivel I	Sentido numérico y pensamiento algebraico	
	Cambios y relaciones	
	Forma, espacio y medida	
	Manejo de la información	

Nivel II	Sentido numérico y pensamiento algebraico	
	Cambios y relaciones	
	Forma, espacio y medida	
	Manejo de la información	

Nivel III	Sentido numérico y pensamiento algebraico	
	Cambios y relaciones	
	Forma, espacio y medida	
	Manejo de la información	

Nivel IV	Sentido numérico y pensamiento algebraico	
	Cambios y relaciones	
	Forma, espacio y medida	
	Manejo de la información	

Formato de puntuaciones de corte
PLANEA EMS 2017

Puntuación de Corte	Número de reactivo Marcador
I	
II	
III	

INTEGRANTES DEL COMITÉ DE NIVELES DE LOGRO Y PUNTUACIONES DE CORTE
EDUCACIÓN MEDIA SUPERIOR

Nombre integrante _____

Nombre integrante _____

Nombre integrante _____

Nombre integrante _____

Nombre integrante _____

Nombre integrante _____

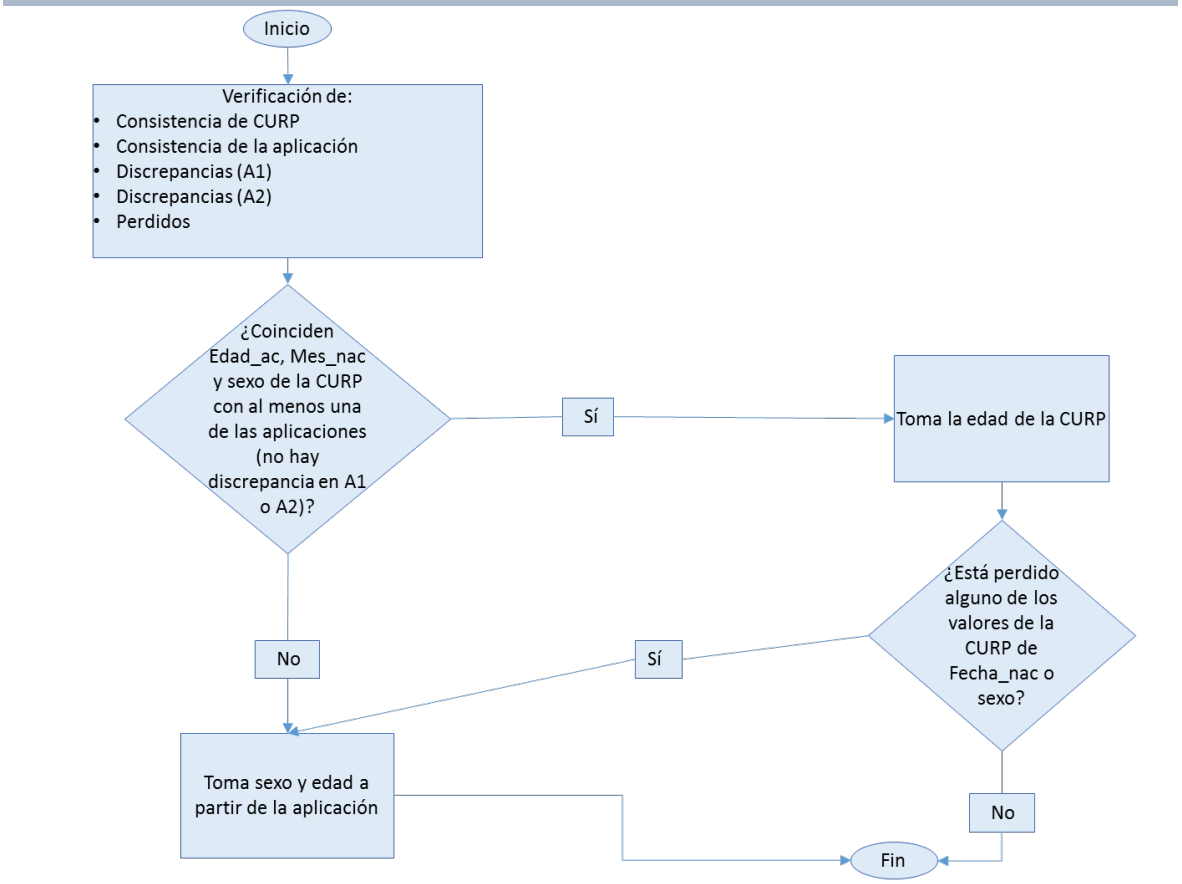
Nombre integrante _____

Anexo XIV. Variables que componen los índices simples y escalas de contexto PLANEA EMS 2017

Las siguientes tablas fueron tomadas del *Reporte técnico de la construcción de índices y escalas de los datos de los cuestionarios de contexto*, de la DGMTD (2018b).

- **Índices simples**

Edad y sexo del alumno



Variables que componen el índice condición indígena

Reactivo	Categorías de respuesta	Codificación
MA_2: ¿Alguno de tus padres habla una lengua indígena? Madre	A) Sí	A) 1
	B) No	B) 2
MA_3: ¿Alguno de tus padres habla una lengua indígena? Padre	C) No sé	C) 3

Construcción del índice condición indígena				
Condición indígena		**MA_2 ¿Alguno de tus padres habla una lengua indígena? Madre**		
		1 Sí	2 No	3 No sé
MA_3 ¿Alguno de tus padres habla una lengua indígena? Padre	1 Sí	1	1	1
	2 No	1	2	9
	3 No sé	1	9	9

Variables que componen los índices nivel de estudios del padre y nivel de estudios de la madre		
Reactivo	**Categorías de respuesta**	**Codificación**
MA_111: ¿Cuál es el nivel de estudios alcanzado por tu padre (contesta aunque ya haya fallecido)? MA_110: ¿Cuál es el nivel de estudios alcanzado por tu madre (contesta aunque ya haya fallecido)?	A) No estudió B) Primaria C) Secundaria D) Nivel medio superior o carrera técnica E) Licenciatura F) Posgrado (especialidad, maestría, doctorado)	A) 1 B) 2 C) 3 D) 4 E) 5 F) 6

Construcción de los índices nivel de estudios del padre y nivel de estudios de la madre		
Índice simple	**Categorías de respuesta**	**Recodificación**
Máximo nivel de estudios del padre Máximo nivel de estudios de la madre	A) No estudió B) Primaria C) Secundaria D) Nivel medio superior o carrera técnica E) Licenciatura o posgrado	A) 1 B) 2 C) 3 D) 4 E) 5

Construcción del índice máximo nivel de estudios alcanzado por los padres						
Máximo nivel de estudios alcanzado por los padres		**MA_111 ¿Cuál es el nivel de estudios alcanzado por tu padre (contesta aunque ya haya fallecido)?**				
		1 No estudió	2 Primaria	3 Secundaria	4 Nivel medio superior o carrera técnica	5 Licenciatura o Posgrado
MA_110 ¿Cuál es el nivel de estudios alcanzado por tu madre (contesta aunque ya haya fallecido)?	1 No estudió	1	2	3	4	5
	2 Primaria	2	2	3	4	5
	3 Secundaria	3	3	3	4	5
	4 Nivel medio superior o carrera técnica	4	4	4	4	5
	5 Licenciatura o Posgrado	5	5	5	5	5

- **Escalas de contexto**

Reactivos que componen la escala Capital económico		
Reactivo	**Reactivo en el cuestionario**	**Etiqueta**
TV	Ma_115 y Ma_119	Índice de televisor: número de aparatos, considerando servicio de cable.
MA_114	Ma_114	Del siguiente listado de bienes, señala cuántos hay en tu casa. Computadora
MA_116	Ma_116	Del siguiente listado de bienes, señala cuántos hay en tu casa. Automóvil
MA_117	Ma_117	En tu casa, ¿cuentas con los siguientes bienes y servicios? Teléfono convencional (fijo)
MA_118	Ma_118	En tu casa, ¿cuentas con los siguientes bienes y servicios? Horno de microondas
MA_120	Ma_120	En tu casa, ¿cuentas con los siguientes bienes y servicios? Internet

Categorías de las variables que componen la escala Capital económico	
Variables	**Categorías**
TV construida a partir de: Ma_115 y Ma_119	0: No tiene televisor, 1: Tiene al menos un televisor (sin servicio de cable o vía satélite), 2: Tiene 1 o 2 televisores y servicio de cable o vía satélite, 3: Tiene 3 o más televisores y servicio de cable o vía satélite
MA_114, MA_116	0: Ninguno, 1: 1, 2: 2 o 3, 3:4 o más
MA _117, MA _118, MA_120	0: No, 1: Sí

<table>
<tr><th colspan="6">Construcción del índice Televisor</th></tr>
<tr><th colspan="2" rowspan="2">(TV) Índice de televisor (número de aparatos, considerando servicio de cable)</th><th colspan="4">MA_119 En tu casa, ¿cuentas con los siguientes bienes y servicios? Televisión por cable o vía satélite</th></tr>
<tr><th>(1) Sí</th><th>(2) No</th><th>(8) Respuesta múltiple</th><th>(9) Respuesta omitida</th></tr>
<tr><td rowspan="6">MA_115 Del siguiente listado de bienes, señala cuántos hay en tu casa. Televisor</td><td>(1) Ninguno</td><td colspan="4">(0) No tiene televisión</td></tr>
<tr><td>(2) 1</td><td rowspan="3">(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3" rowspan="3">(1) Tiene al menos un televisor</td></tr>
<tr><td>(3) 2</td></tr>
<tr><td>(4) 3</td></tr>
<tr><td>(5) 4 o más</td><td>(3) Tiene 3 o más televisores y servicio de cable o vía satélite</td></tr>
<tr><td>(8) Respuesta múltiple</td><td>(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3">( ) Respuesta omitida</td></tr>
</table>

ANEXOS TÉCNICOS

Anexo 1. Diseño muestral

En el ciclo escolar 2014-2015 inició la implementación del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA), el cual evalúa a los alumnos para medir el dominio de aprendizajes considerados clave en la educación obligatoria.

En el ciclo 2016-2017, el Instituto Nacional para la Evaluación de la Educación (INEE) fue responsable de diseñar y aplicar las pruebas correspondientes a la Evaluación de Logro referida al Sistema Educativo Nacional (ELSEN) para los alumnos del grado terminal de educación media superior, mientras que la Secretaría de Educación Pública (SEP) aplicó las pruebas correspondientes a la Evaluación del Logro referida a los Centros Escolares (ELCE). La ELSEN se aplicó a una muestra representativa de alumnos, con el fin de contar con estimaciones confiables a nivel nacional, por tipo de escuela y por entidad federativa, mientras que la ELCE se aplicó a una muestra de alumnos de cada una de las secundarias no seleccionadas en la muestra de la ELSEN.

El propósito del presente documento es recopilar la información referente al diseño de la muestra seleccionada para la aplicación de la prueba PLANEA en el nivel de educación media superior, desde la planeación hasta la selección de la muestra final, incluyendo las diversas modificaciones llevadas a cabo en el transcurso que resultaron en cambios en la planeación.

1.1 Población objetivo

La implementación 2017, tanto de la Elsen como de la Elce, se enfoca en la evaluación de los alumnos del año o grado terminal de educación media superior (ems). A diferencia de preescolar, primaria y secundaria, los programas y carreras de este tipo educativo no tienen una duración única fija, por lo que es posible encontrar centros de trabajo cuyos programas duran entre 2 y 5 años. Incluso dentro del mismo centro escolar se pueden encontrar diferentes programas con distinta duración, lo que conlleva que la población objetivo esté compuesta por una mezcla de alumnos de diferentes grados.

1.2. Dominios de estudio

Los dominios de estudio son subpoblaciones de interés sobre las que se quiere obtener y comunicar resultados con una precisión específica. Para la implementación de PLANEA EMS 2017, los dominios de estudio definidos, y sobre los que se darán resultados, son los siguientes:

- *Nacional*: Todos los alumnos cursando el grado terminal de EMS en el país.
- *Entidad*: Alumnos en el grado terminal de EMS en cada una de las 32 entidades del país.
- *Sostenimiento*: Alumnos en el grado terminal de EMS, de cada uno de los cuatro tipos de sostenimiento de EMS a nivel nacional (Autónomo, Estatal, Federal, Privado).
- Principales subsistemas: Alumnos en el último grado de los ocho subsistemas nacionales que acumulan la mayor proporción de alumnos de la población objetivo. Estos son:

1. Particulares
2. Colegio de Bachilleres
3. DGETI
4. Bachillerato estatal DGE-CGE
5. Bachillerato autónomo
6. CECYTE
7. CONALEP estatal
8. Telebachilleratos

- *Entidad-Sostenimiento*: Alumnos que cursan el último grado de EMS en cada uno de los cuatro tipos de sostenimiento en cada una de las treinta y dos entidades federativas del país. Es necesario aclarar que, como se mostrará más adelante en el documento, no será posible obtener estimaciones lo suficientemente confiables en todas las posibles combinaciones de sostenimiento-entidad que existen dada la población en dichas combinaciones.

1.3. Integración inicial del marco muestral

1.3.1 Fuentes de información que conforman el marco muestral

Para la planeación de esta aplicación de PLANEA se recurrió a dos fuentes principales de información sobre la matrícula actual de alumnos en el grado terminal.

La primera de ellas fue la Estadística Oficial: Información actualizada al inicio del ciclo 2016-2017, registrada en los centros de trabajo mediante el Formato Estadístico 911, que fue proporcionada a la Subdirección de Muestreo del INEE por la Subsecretaría de Educación Media Superior. La matrícula contenía, además del total de alumnos por modalidad (mixta o escolarizada) según el grado, información sobre los turnos impartidos en cada modalidad, pero no la matrícula por turno. La información contenida en la variable "Duración" (del programa de estudio) no fue del todo confiable.

La segunda fuente fue la información enviada voluntariamente por los centros de trabajo a la Secretaría de Educación Pública (SEP), en la que se reporta directamente el listado de los alumnos matriculados en el grado terminal hasta los primeros días de enero del año en curso. Es pertinente mencionar que para algunos centros de trabajo el listado con los nombres de los alumnos no está disponible, pero sí la cantidad de alumnos del grado terminal.

1.3.2 Planteles y escuelas

De acuerdo con la Subsecretaría de Educación Media Superior (SEMS), *plantel* es un conjunto organizado de recursos humanos y físicos que, bajo la autoridad de un director o responsable, desarrolla actividades del proceso de enseñanza-aprendizaje en el tipo educativo medio-superior. Un plantel realiza las actividades educativas en uno o más turnos en la instalación sede y en algunas ocasiones también lo hace en una o varias extensiones.

Se define como *escuela* a la combinación de la Clave del Centro de Trabajo, la extensión y el turno impartido.

El marco inicial de escuelas, con su respectiva matrícula, se compone de la información enviada por la SEP, así como de la información de la base proporcionada por la Subsecretaría relacionada con aquellos centros de trabajo cuyos datos no fueron encontrados en la base de la SEP.

Tomando como base la información proporcionada por la SEP, es posible identificar las escuelas y calcular su matrícula, ya que se cuenta con toda la información necesaria. Sin embargo, en el caso de la información de la Subsecretaría, se desconoce qué centros de trabajo son o tienen extensiones y no se dispone de una cifra exacta de la matrícula de cada turno impartido. Por lo tanto, las escuelas identificadas en la base de datos de la Estadística Oficial son únicamente la combinación de la clave del Centro de Trabajo y el turno impartido.

1.3.3 Duración y matrícula de las escuelas de la Estadística Oficial

Dado que la información de la Estadística Oficial presentó inconsistencias entre la duración reportada y los grados con matrícula reportada, se utilizó una serie de criterios para determinar qué matrícula considerar como la del año terminal.

Para poder considerar a las escuelas no incluidas en el marco de la SEP se determinó, además, una aproximación de la matrícula por turno a partir del marco de EMS del ciclo 2015-2016, de manera que los centros de trabajo con más de un turno recibieron la misma proporción de matrícula reportada en cada turno en el ciclo anterior. Para los centros de trabajo que no reportaron más de un turno en el ciclo anterior, o que no fueron encontrados en la base de datos, pero que en la Estadística Oficial sí tienen más de un turno, se distribuyó su matrícula con el promedio nacional de la proporción por turno de las escuelas que impartieron esos turnos en el marco anterior. Por ejemplo, en una escuela con turnos matutino y nocturno en modalidad escolarizada se tomó el promedio de las proporciones por turno de todas las escuelas que imparten únicamente estos dos turnos en modalidad escolarizada, sin tomar en cuenta las escuelas que tuvieron matrícula cero en alguno de los turnos.

1.3.4 Turnos

En cada Centro de Trabajo con turno discontinuo se determinó que la aplicación de la prueba se llevara a cabo en la mañana. Adicionalmente se tomó la decisión de combinar los turnos matutino y discontinuo cuando ambos fueran impartidos. Por esta razón en el marco de muestreo, el turno discontinuo se recodificó como matutino y cuando el turno matutino también tenía matrícula reportada, se sumaron ambas matrículas y se reportó como turno matutino. Ello implica que, de realizarse cualquier tipo de análisis sobre los turnos, el reporte del turno matutino será en realidad sobre los alumnos del turno matutino y discontinuo.

1.4 Exclusiones

En el marco muestral se excluyeron:

a) Escuelas que contienen el grupo de letras 'PET' al substraer los elementos cuatro, cinco y seis de la Clave del Centro de Trabajo, ya que son escuelas de capacitación técnico-profesional.
b) Escuelas identificadas como penales o militares.

c) Escuelas con matrícula nula en el último grado.
d) Escuelas con duración menor a dos años.
e) La matrícula total aproximada de alumnos en el país que se encuentran en el grado terminal de EMS es 1 316 050. Eliminando la matrícula de las escuelas registradas como penales o militares, la matrícula final considerada es 1 315 116, por lo que 99.93% de la población objetivo se encuentra incluida en el marco final. Es importante mencionar que estos totales se obtuvieron de la última actualización del marco muestral, explicada en la sección 17. La exclusión de los casos indicados en los incisos a y d se debe a que no forman parte de la población objetivo de PLANEA EMS 2017.

1.5 Estratos

Los estratos están compuestos por el cruce entre las entidades, los tipos de sostenimiento y los subsistemas que conforman cada uno de ellos. Esto significó la definición de 414 estratos.

	Subsistemas del sostenimiento Autónomo	Subsistemas del sostenimiento Estatal	Subsistemas del sostenimiento Federal	Subsistemas del sostenimiento Privado
Entidades	Estratos			

1.6 Método de selección de la muestra

El esquema de muestreo propuesto es bietápico.

Primera etapa: Para la selección de las Unidades Primarias de Muestreo (UPM) se propuso un esquema de Probabilidad Proporcional al Tamaño Sistemático (PPT), en cada dominio de estudio.

Segunda etapa: Para la selección de las Unidades Secundarias de Muestreo (USM) se estableció un esquema de Muestreo Aleatorio Simple en cada UPM. En esta etapa se delimitaron las cuotas de tamaño muestral para cada escuela de la siguiente manera:

i. Si la matrícula del grado terminal es menor o igual que 40, se censó a la escuela.
ii. Si la matrícula del grado terminal es mayor que 40 y menor que 80, se seleccionó una muestra de 40 alumnos.
iii. Si la matrícula del grado terminal es mayor o igual que 80, se seleccionaron 80 alumnos.

Adicionalmente, fue necesario obtener una submuestra de la muestra original de escuelas en las que se seleccionarían algunos alumnos para participar en la aplicación de las pruebas que van a servir para la 'Comparabilidad' de las escalas de puntajes entre las aplicaciones de PLANEA EMS 2015, 2016 y 2017.

Durante la planeación inicial, las UPM consideradas fueron las escuelas, y las USM los alumnos. Con base en esta definición, se realizó el cálculo del tamaño muestral para cada dominio de estudio establecido. Posteriormente el diseño muestral sufrió modificaciones, por lo que las UPM fueron redefinidas, como se explica en la sección 13.

1.7 Precisión de los estimadores

La precisión se define como el inverso de la varianza de un estimador (Kish, 1965: 25). Bajo un esquema de muestreo aleatorio simple, el estimador de la varianza de la media muestral de una variable Y, $\bar{y}$, tiene la expresión:

$$S_{\bar{y}}^2 = \left(1 - \frac{n}{N}\right)\frac{S_y^2}{n},$$

donde S_y^2 es la varianza real pero desconocida de la variable Y. Mientras más pequeño sea el valor de $S_{\bar{y}}^2$, mayor será su precisión, γ. Definiendo:

$$\gamma = \frac{1}{c^2 S_y^2}, \;\; c > 0,$$

es posible obtener un valor de n para diferentes valores fijos de c mediante la relación:

$$n = \frac{1}{c^2 + \frac{1}{N}} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Lo anterior implica tres puntos importantes:

i. Mientras más pequeño sea el valor de c, mayor será el valor de n.
ii. Mientras mayor sea el valor de n, mayor será la precisión.
iii. La precisión puede ser interpretada a partir del valor de c. Mayor precisión será obtenida si el valor de c es más pequeño.

Debido a la relación entre la precisión y la varianza del estimador, la precisión también afecta los intervalos de confianza calculados para el estimador. Bajo supuestos de normalidad, la expresión de un intervalo de confianza del (1-α) x 100% para el estimador $\bar{y}$ es de la forma:

$$\bar{y} \pm z_{1-\frac{\alpha}{2}}\, S_{\bar{y}},$$

donde $z_{1-\frac{\alpha}{2}}$ representa el cuantil correspondiente de la distribución normal estándar, es decir, $z_{1-\frac{\alpha}{2}} \coloneqq \{x: P(X \leq x) = 1 - \alpha/2\}$ donde $X \sim N(0,1)$.

Para crear un nivel de confianza de 95% suele utilizarse la aproximación $z_{1-\frac{\alpha}{2}} \approx 2$, por lo que el intervalo de confianza tiene la expresión:

$$\bar{y} \pm 2\, S_{\bar{y}},$$

o equivalentemente:

$$\bar{y} \pm 2\, cS_y.$$

En la expresión anterior puede observarse que el valor de c está directamente asociado con el margen de error de $\bar{y}$, por lo que valores menores de c conllevarán intervalos de confianza de menor amplitud.

En estudios como la edición del 2003 de TIMSS, la del 2004 de CivEd y la del 2006 de PIRLS establecen como estándar un tamaño muestral efectivo de al menos 400, lo cual conlleva a un valor de $c = 0.05$.

Considerando este valor de c, la expresión del intervalo de confianza es igual a $\bar{y} \pm 0.1\, S_y$, la cual es reportada en estos estudios de la IEA como el intervalo de confianza aproximado de 95% para la media, utilizando una estimación de S_y al ser ésta desconocida. En la aplicación de TIMSS y PIRLS del 2011 el valor de c es fijado en un valor menor, de 0.035.

1.8 Tamaños muestrales planeados para cada dominio de estudio

1.8.1 Determinación del tamaño de muestra efectivo para las entidades, Sostenimientos y combinación entidad-sostenimiento

Las precisiones finales aproximadas se presentan en la sección 12, al final del esquema planeado. Las secciones que siguen describen los pasos mediante los cuales se obtuvieron estas precisiones.

El tamaño de muestra efectivo de alumnos necesario a nivel estatal, n_E, fue calculado utilizando la expresión en (1), fijando inicialmente el valor de $c = 0.07$. Para un intervalo de confianza de 95% este valor de c implica un margen de error aproximado de 14% de la estimación de la desviación estándar. En un principio, el valor de c fue fijado en 0.05; sin embargo, por razones de presupuesto fue modificado para que el tamaño muestral de escuelas disminuyera.

Posteriormente, n_E fue dividido en los sostenimientos existentes en la entidad de manera proporcional a la matrícula de cada sostenimiento, para obtener el tamaño de muestra efectivo a nivel entidad-sostenimiento, n_{ES}.

Para cada valor n_{ES} es posible obtener el valor de c correspondiente, así como el margen de error equivalente, mediante

$$c = \sqrt{\frac{1}{n_{ES}} - \frac{1}{N_{ES}}},$$

donde N_{ES} es la matrícula total de alumnos en el grado terminal en el sostenimiento S de la entidad E.

Dado que las precisiones obtenidas no resultaron altas (es decir, que se tuvieron valores de c grandes), se procedió a obtener un nuevo valor de n_{ES}, denominado por n_{ij}^0, utilizando la siguiente regla para el valor c:

i. Si el margen de error aproximado para un intervalo de 95% de confianza era mayor o igual a 24% de la desviación estándar ($c \geq 0.12$), el margen de error se fijó en 24% de la desviación estándar.
   En otro caso, el margen de error aproximado para el estrato conformado por la i-ésima entidad y el j-ésimo sostenimiento se fijó considerando el promedio de los márgenes de error de los estratos dentro de la misma entidad que no sean mayores a 24% de la desviación estándar.

Cabe mencionar que en Jalisco la regla fue aplicada si $c \geq 0.13$, y en San Luis Potosí la regla se aplicó si $c \geq 0.11$. Las aproximaciones de los márgenes de error fueron fijados en 26% y 22% de la desviación estándar, respectivamente.

Partiendo de los márgenes de error fijados, y sus respectivos valores de c, se calculó el valor n_{ij}^0 con la expresión (1). La suma de los tamaños de muestra efectivo en cada combinación ij sobre todos los sostenimientos dados por j, es el nuevo tamaño de muestra efectiva a nivel estatal, n_i^0. A partir de n_i^0 es posible calcular el valor de c que le corresponde, así como el margen de error aproximado utilizando la relación

$$c = \sqrt{\frac{1}{n_i^0} - \frac{1}{N_i}}, \quad i = 1,2,3, \dots ,32,$$

donde N_i representa el total de alumnos en el grado terminal de la i-ésima entidad.

La siguiente tabla ejemplifica el proceso anterior.

Entidad	Sostenimiento	Valor de c	Valor de c deseado		Tamaño de muestra efectivo	Tamaño de muestra efectivo a nivel estatal	Valor de c para la media a nivel estatal
i	Autónomo	0.31	.	0.12	65	302	0.057
	Estatal	0.10	0.11	0.11	85		
	Federal	0.12	0.11	0.11	84		
	Particular	0.21	.	0.12	68		

1.8.2 Efecto del diseño

Grupos de cuotas y cuotas

Las escuelas incluidas en el marco fueron clasificadas en tres grupos de cuota a partir de la matrícula estimada o reportada en cada una de ellas con el fin de identificarlas como escuelas pequeñas, medianas o grandes, según la cantidad de alumnos que podrían participar en la prueba PLANEA. Debido a que algunas escuelas serán elegidas para la aplicación de las pruebas que van a servir para la 'Comparabilidad' de las escalas de puntajes entre las aplicaciones de PLANEA EMS 2015, 2016 y 2017 en una porción de los alumnos elegidos en la muestra, la cantidad de alumnos seleccionados que contestarán la prueba PLANEA se verá reducida porque cada alumno sólo puede participar en una aplicación.

Como aún no se había seleccionado la submuestra de escuelas destinadas a la aplicación de las pruebas de 'Comparabilidad', se ajustó la matrícula original en todas las escuelas siguiendo este criterio:

Matrícula original	Matrícula ajustada
Menor que 5	Matrícula
5 a 9	Matrícula - 1
10 a 39	Matrícula - 2
40 a 75	38
76 o más	76

El ajuste a la matrícula se hizo considerando que en todas las escuelas de la muestra con más de 5 alumnos matriculados se obtendrá una submuestra para la 'Comparabilidad', en la que el número de alumnos seleccionados varía según la matrícula de la escuela. Considerando la matrícula original se definieron los grupos de cuota; asimismo, se establecieron las cuotas según la matrícula ajustada de la siguiente manera:

Estrato*	Grupo de cuota	Matrícula original	Cuota
	1	Menor que 40	Promedio de la matrícula ajustada de las escuelas en el grupo
	2	40 a 75	38
	3	Mayor o igual a 76	76

*****Conformado por la i-ésima entidad y el j-ésimo sostenimiento, $i = 1,2,\dots,32$, $j = 1,2,3,4$.

Estimación del efecto del diseño ($deff$).

Sea q_k la cuota en cada escuela de acuerdo al grupo de cuota k. Considerando las cuotas establecidas según el tamaño de la escuela, y que se requieren entre uno y cuatro alumnos por escuela submuestreada para la aplicación de las pruebas de 'Comparabilidad', se estimó el $deff$ como se describe a continuación:

$$deff_{ijk} = 1 + (q_k - 1)\rho$$

donde $q_k = \overline{M}_{ij}, 38, 76$**,** $k = 1,2,3$, y el valor de $\rho = 0.18$**.**

1.8.3 Tamaño de muestra preliminar considerando el efecto del diseño

Sea n_{ijk}^0 el tamaño de muestra efectivo n_{ij}^0 dividido proporcionalmente a la matrícula de todas las escuelas clasificadas en el k-ésimo grupo de cuota, dentro de los 3 grupos especificados en la sección 9.2.1.

El tamaño muestral (no efectivo) preliminar de cada j-ésimo sostenimiento en la i-ésima entidad, n_{ij}^p, se obtuvo con la siguiente relación considerando una tasa de no respuesta del 5%, es decir, $TNR = 0.05$:

$$n_{ij}^p = \sum_{k=1}^{3} \frac{n_{ijk}^0 * deff_{ijk}}{1 - TNR}\,, \qquad i = 1,2,\dots,32, \qquad j = 1,2,3,4.$$

Así, el tamaño muestral preliminar de cada entidad es

$$n_i^p = \sum_{j=1}^{4} n_{ij}^p\,, \qquad i = 1,2,\dots,32, \qquad j = 1,2,3,4.$$

1.8.4 Número de escuelas y tamaño de muestra preliminar corregido

Para conocer el número de escuelas necesarias para cubrir el tamaño n_{ij}^{p} se considera

$$e_{ijk} = \frac{n_{ijk}^{0} * deff_{ijk}}{(1 - TNR)q_k}, \qquad i = 1,2, \dots ,32, \qquad j = 1,2,3,4, \qquad k = 1,2,3,$$

donde k representa el grupo de cuota, y q_k la cuota correspondiente.

Sea er_{ijk} el número real de escuelas existentes en el k-ésimo grupo de cuota del j-ésimo sostenimiento en la i-ésima entidad. Cuando $e_{ijk} > er_{ijk}$, se hace necesario definir el número de escuelas corregido de la siguiente manera:

$$ec_{ijk} = \begin{cases} e_{ijk}\ si\ e_{ijk} \leq er_{ijk} \\ er_{ijk}\ si\ e_{ijk} > er_{ijk}, \end{cases}$$

por lo que el número de alumnos del estrato ijk por muestrear es

$$nc_{ijk} = \begin{cases} \dfrac{n_{ijk}^{0} * deff_{ijk}}{1 - TNR} & si\ e_{ijk} \leq er_{ijk}, \\ ec_{ijk} * q_k & si\ e_{ijk} > er_{ijk} \end{cases}$$

de manera que $nc_{ij}^{p} = \sum_{k=1}^{3} nc_{ijk}$ es el tamaño muestral de alumnos corregido.

1.9 Asignación de la muestra considerando subsistemas

Los principales subsistemas de educación media superior en el país forman parte de los dominios de estudio. Para garantizar la inclusión de todos los subsistemas en la muestra, el subsistema fue considerado como variable de estratificación junto con la entidad y el sostenimiento.

La distribución del tamaño muestral nc_{ij}^{p} a través de los diferentes subsistemas presentes en el j-ésimo sostenimiento de la i-ésima entidad surge de manera natural como proporcional a la matrícula que representa cada subsistema en tal sostenimiento y entidad. Sin embargo, distribuir la cantidad de alumnos conllevaría aumentar el número de escuelas a evaluar, ya que se tendría un menor número de alumnos a evaluar en escuelas de todos los subsistemas. Por ello se realizó la distribución tomando las proporciones de escuelas de cada intersección entre subsistema, sostenimiento y entidad considerando los grupos de cuota, denotadas como p_{ijkl}, y redistribuyendo el número de escuelas en vez de la matrícula.

Sea ef_{ijkl} el número final de escuelas necesarias en la muestra que sean del l-ésimo subsistema del k-ésimo grupo de cuota del j-ésimo sostenimiento en la i-ésima entidad. Para recalcular el número de alumnos en cada subsistema según los grupos de cuotas, denotado como nf_{ijkl}, se multiplicó la cuota asociada al k-ésimo grupo de cuota por ef_{ijkl}, es decir

$$nf_{ijkl} = ef_{ijkl} * q_k.$$

La tabla siguiente muestra un ejemplo de dicho proceso.

Entidad	Sostenimiento	Subsistema	Grupo de cuota	Cuota			Número de escuelas	Total de escuelas por grupo de cuota				
Aguascalientes	Federal	DGETI	1	35	1	4	1	1	1.000	1	35	35
		DGB	2	38	2	43	1	7	0.143	0	0	38
		DGECYTM					1		0.143	0	0	
		DGETI					5		0.714	1	38	
		DGB	3	76	14	965	5	26	0.192	3	228	1140
		DGECYTM					1		0.038	1	76	
		DGETA					6		0.231	3	228	
		DGETI					14		0.538	8	608	

1.10 Sobremuestra en la Ciudad de México

Previniendo una posible falta de participación de las escuelas en los subsistemas de la Universidad Nacional Autónoma de México (UNAM) y del Instituto de Educación Media Superior (IEMS) en la Ciudad de México, se determinó incrementar el tamaño muestral en la entidad calculando dos tamaños de muestra tal como se especificó en las secciones 9 y 10, pero tomando diferentes marcos de muestreo para esta entidad. El primer marco es el mismo descrito previamente, mientras que el segundo elimina la información de los dos subsistemas mencionados.

Sea nf_{ijkl}^1 el tamaño muestral obtenido utilizando el marco completo y nf_{ijkl}^2 el tamaño muestral calculado con el marco reducido. Entonces, el tamaño de muestra final para la Ciudad de México es igual a nf_{ijkl}^1, cuando el subsistema corresponde a la UNAM o al IEMS, y es igual a $\max(nf_{ijkl}^1, nf_{ijkl}^2)$ en otro caso. Los tamaños muestrales finales para los demas estratos no sufrieron modificación alguna.

1.11 Márgenes de error basados en el esquema planeado

El cambio en el tamaño muestral, nf_{ijk}, respecto del propuesto, nc_{ijk}, implica un cambio también en la precisión estimada en cada uno de los dominios de estudio. Los márgenes de error aproximados después de la modificación se resumen de la siguiente manera:

Dominio	Márgenes de error aproximados	c
Nacional	2% de S_y aproximadamente	0.01
Entidad	<12.8% de S_y	
Sostenimiento	<5.4% de S_y	
Principales subsistemas a nivel nacional	< 10.2% de S_y, excepto Telebachilleratos donde el márgen de error aproximado es < 14% de S_y	
Entidad-Sostenimiento	< 28% de S_y en la mayoría de los casos**	

** Para los siguientes casos los márgenes de error son mayores que el 28% de S_y:

Entidad	Sostenimiento
Querétaro	Autónomo
Colima	Federal
Campeche	Autónomo
Durango	Autónomo
Tamaulipas	Autónomo
Aguascalientes	Autónomo
Yucatán	Autónomo
San luis potosí	Autónomo

1.12 Modificación del diseño muestral

Como consecuencia de los acuerdos entre el INEE y la SEP para la homologación del proceso de aplicación de la prueba ELSEN con la aplicación de la prueba ELCE, se determinó un cambio en la definición de las UPM. Éstas pasaron de ser las escuelas a definirse como la concatenación de los Centros de Trabajo y extensión (CCT+Extensión).

Para cada UPM seleccionada en la muestra, se censarán los turnos impartidos en ella, así como los grupos escolares existentes en cada turno.

Los estratos de la segunda etapa fueron conformados por los grupos escolares de cada turno impartido en cada UPM cuando esta información estuvo disponible. En las UPM sin información relacionada con los grupos únicamente se estratificó por turno. Así, la segunda etapa de muestreo fue modificada a un esquema Estratificado, realizando Muestreo Aleatorio Simple en cada estrato.

1.13 Modificación del tamaño muestral final de UPM

Considerando el cambio en la definición de las UPM se reajustó el tamaño muestral de escuelas para obtener el tamaño muestral de Centros de Trabajo-Extensión a utilizar. Sea nf_{ijkl}^{UPM} el tamaño muestral a determinar de las UPM modificadas y $r(x), x \geq 0$, la función de redondeo estándar. Entonces,

$$nf_{ijkl}^{UPM} = \begin{cases} r(p_{ijkl}^2 * nf_{ijkl}) \; si \; r(p_{ijkl}^2 * nf_{ijkl}) > 0 \\ 1 \qquad\qquad\qquad\qquad\qquad en \; otro \; caso \end{cases}.$$

1.14 Selección de la muestra de la primera etapa (UPM)

Previamente a la selección de la muestra de UPM, el marco muestral de las mismas fue ordenado mediante el método *Serpentine* utilizando las variables:

- Entidad
- Sostenimiento
- Subsistema
- Matrícula

Las UPM fueron seleccionadas en la primera etapa del muestreo, mediante selección de PPT Sistemático en cada uno de los estratos definidos.

1.15 Selección de la submuestra de escuelas

La selección de las escuelas participantes en las pruebas de 'Comparabilidad', se realizó mediante Muestreo Aleatorio Simple de las escuelas que conforman las UPM seleccionadas en la muestra de la primera etapa.

En cada escuela seleccionada en la submuestra se tomaron 1, 2 o 4 alumnos, de acuerdo con el siguiente criterio basado en la matrícula del grado terminal:

- Si la matrícula de la escuela es menor que 2, entonces se realizó un censo de la matrícula. En las escuelas con esta característica no se aplicará la prueba de PLANEA, únicamente la prueba de 'Comparabilidad'.
- Si la matrícula de la escuela es mayor que 2 pero menor que 80, se asignaron 2 alumnos a la aplicación de la prueba de 'Comparabilidad' y el resto de la cuota a la prueba PLANEA.
- Si la matrícula de la escuela es mayor o igual que 80, 4 alumnos se destinaron a la aplicación de la prueba de 'Comparabilidad' y el resto de la cuota a la prueba PLANEA.

1.16 Actualización del marco de centros de trabajo y el marco de alumnos

Posterior a la selección de la muestra de la primera etapa del esquema propuesto, la SEP proporcionó información actualizada sobre ocho centros de trabajo cuyos datos fueron obtenidos previamente mediante la Estadística Oficial y que fueron seleccionados en la muestra. Debido a que algunos de ellos reportaron diferencias en la matrícula registrada, se procedió a incorporar la nueva información tanto en el marco de centros de trabajo como en el de alumnos, pero como la muestra de las UPM ya se había realizado, la modificación de los datos no se refleja en la selección de la primera etapa. Sin embargo, sí se puede reflejar en la estimación de los márgenes de error que se espera observar, como se explica en la sección 19.

En el caso de la muestra de la segunda etapa, los cambios en el marco de alumnos sí estuvieron incorporados en las probabilidades de selección de primer orden, siguiendo las modificaciones en el esquema de esta etapa, descritas en la sección 13.

1.17 Tamaño de muestra al interior de las UPM seleccionadas

1.17.1 Asignación de la muestra

Dado que la m-ésima UPM fue seleccionada en la primera etapa del esquema de muestreo, se determinó el tamaño de muestra de alumnos en cada uno de los turnos que imparte a través de las cuotas especificadas en la sección 9.3 aplicadas al marco de alumnos actualizado.

Asimismo, la cuota de cada escuela fue distribuida de manera proporcional a la matrícula de cada grupo reportado en ella. Cuando esta información no se encontró, se consideró la matrícula como un único grupo del cual se eligiría la cuota correspondiente.

1.17.2 Selección de la muestra de alumnos

Los alumnos fueron seleccionados en la segunda etapa del muestreo utilizando Muestreo Aleatorio Simple en cada estrato definido por la intersección entre la Escuela y los Grupos en ella.

Para la construcción del marco muestral de alumnos se utilizó la versión actualizada de la base original, según se explicó en la sección 17.

1.18 Modificación de las características de la muestra seleccionada

Derivado del cambio en la definición de las UPM y de la consecuente modificación del tamaño muestral, así como la incorporación de la información actualizada sobre ocho UPM en el marco muestral, se dieron variaciones en la precisión estimada y el correspondiente valor de c a partir de la muestra obtenida. A continuación se describe la estimación considerando las modificaciones descritas, la muestra obtenida en la primera etapa y la submuestra ya seleccionada.

Para cada dominio de estudio, d, se tiene un número n_d de alumnos seleccionados en la muestra. Para realizar el despeje de la expresión en (1) es necesario conocer el tamaño de muestra efectivo del dominio, denotado por n_d^0, el cual es aproximado mediante la relación

$$\hat{n}_d^0 = \frac{n_d}{deff_d} * (1 - TNR_d),$$

siendo $deff_d$ un estimado del efecto del diseño, y TNR_d la Tasa de No Respuesta, del dominio d. La TNR_d considerada fue 0.05 para cualquier dominio d.

Sean
U_d ≔ Total de UPM en el dominio d,
ε_m ≔ Número de escuelas en la m-ésima UPM, $m = 1,2, \dots, U_d$,
q_p ≔ Cuota de la p-ésima escuela de la m-ésima UPM como se especificó en la sección 7, $p = 1,2, \dots, \varepsilon_m$ y
t_p ≔ Número de alumnos en la p-ésima escuela destinados a la aplicación de las pruebas de 'Comparabilidad'. Si la escuela no fue seleccionada en la submuestra, $t_p = 0$.
Definiendo

$$q_p^* = q_p - t_p, \qquad \text{y}$$
$$\bar{s}_d = \frac{\sum_{m=1}^{U_d} \sum_{p=1}^{\varepsilon_m} q_p^*}{U_d},$$

el estimador propuesto para el efecto del diseño de cada dominio d es

$$deff_d = 1 + (\bar{s}_d - 1) * 0.18.$$

Una vez obtenido $\hat{n}_d^0$ se puede estimar el valor de c del dominio como

$$c_d = \sqrt{\frac{1}{\hat{n}_d^0} - \frac{1}{N_d}},$$

donde N_d es el total de alumnos inscritos en el dominio d.

Las tablas que resumen los valores de c estimados a partir de la muestra seleccionada para los diferentes dominios de estudio se encuentran en el anexo II.

1.19 Ponderación

El propósito de utilizar los factores de expansión o ponderadores (Kalton, 1983: 69; Särndal, 1992: 42; Sharon, 1999: 223-227) sobre los archivos de datos para las evaluaciones a gran escala y otras encuestas es ayudar a los usuarios de los datos a obtener estimaciones de parámetros poblacionales que:

- No sufran sesgo resultante del uso de un diseño muestral complejo.
- Tengan un sesgo mínimo debido a la presencia de no respuesta en la medida de lo posible.

En la mayoría de las encuestas hay una serie de pasos que se realizan durante el proceso de ponderación, éstos incluyen el cálculo de los pesos base (también conocidos como pesos de diseño), ajustes por no respuesta y algún método para reducir la varianza.

En muestras probabilísticas, el peso base es el producto del inverso de las probabilidades de selección de la unidad en cada etapa de muestreo. Dichos pesos pueden ser usados para ponderar una muestra a la población finita, siempre y cuando el marco sea perfecto y todas las unidades respondan. Sin embargo, no todas las unidades responden ya sea porque al momento de la aplicación algunas unidades son no elegibles, no existen o no son de interés, o son elegibles pero no participan por distintas razones. Sin ajuste por no respuesta puede haber niveles de sesgo significativo en las estimaciones; de ahí la necesidad de ajustar los pesos base para conseguir los pesos finales.

El ajuste de los pesos base es un paso indispensable que a su vez puede ocasionar una asignación grande de pesos a ciertas unidades. Algunas de las razones por las que se puede encontrar pesos desiguales es: 1) el método de muestreo, por ejemplo, en muestreo PPT las tasas dependen del tamaño de una variable auxiliar; 2) el número de elementos de la población en los estratos; 3) las tasas de respuesta distintas en los estratos ocasionando ajustes desiguales y; 4) la actualización de la información en los marcos (Potter, 1990: 7-10).

La presencia de pesos grandes puede resultar en grandes varianzas de muestreo en las estimaciones puntuales de interés, incluso si éstos reducen el sesgo (Potter, 1990: 1-2). A menudo se utilizan procedimientos para recortar los pesos, especialmente los grandes (Valliant, Dever, & Kreuter, 2013: 388; Potter, 1990). En la práctica se utilizan métodos *ad hoc*, por ejemplo, en National Assessment of Educational Progress (NAEP) y Programme for International Student Assessment (PISA) aplican el denominado *trimming*, donde los pesos grandes son ajustados (recortados) a una cota y el total de los pesos recortados se reparte entre las otras unidades de muestreo.

El cálculo de los ponderadores debe contemplar todas las etapas hasta llegar a las unidades últimas de muestreo (UUM), además de todos los aspectos mencionados anteriormente y que generan variaciones en los pesos, por lo que de aquí en adelante se describen las consideraciones y pasos para llegar al peso final de los estudiantes.

1.20 Elementos de la estimación del peso

Las notaciones y elementos que se utilizan en este documento y que se refieren a las unidades de muestreo son las siguientes:

- Se consideraron H estratos explícitos; indexados por $h = 1, \dots, H$.
- Se extrajo una muestra S^hde m_h UPMs indexadas por $i = 1, \dots, m_h$ de una población U_h de M_h elementos que comprenden el estrato h.
- Cada UPM i que forma parte de la muestra se registró con t_i turnos, de donde se tomaron todos para la aplicación.
- En cada turno $j = 1, \dots, t_i$ de la UPM i, se estratificó por grupos escolares existentes para seleccionar n_{ij}^h alumnos, indexados por $l = 1, \dots, n_{ij}^h$, de N_{ij}^h disponibles.

∗ El número de alumnos en la UPM $i = 1, \dots, M_h$ del estrato h es $N_{hi} = \sum_{j=1}^{T_i} N_{ij}^h$, por lo tanto, el número de alumnos en la población dentro del estrato $h = 1, \dots, H$ está dado por $\sum_{i=1}^{M_h} N_{hi}$.

En el esquema PPT, en este caso proporcional a la matrícula, las probabilidades deben satisfacer la regla de proporcionalidad respecto a la matrícula ($\pi_i \propto N_{hi}$), pero, debido a las grandes variaciones en la matrícula en grado terminal de las UPMs, no fue posible satisfacer estrictamente la regla. Para lidiar con este problema se definieron las probabilidades de selección de tal forma que las UPM de tamaño muy grande tuvieron probabilidad 1 de ser seleccionadas y de esta forma el resto recupera la regla de proporcionalidad (Särndal, Swensson, & Wretman, 1992): 89-90). A estas unidades se les conoce como unidades de inclusión forzosa o de certeza.

Definimos A_h al conjunto de m_{h_A} UPM de inclusión forzosa en cada estrato h, es decir, el conjunto de las UPM que satisfacen $m_h N_{hi} > \sum_{i=1}^{M_h} N_{hi}$; en dicho conjunto el peso base de las UPM es 1.

Otra consideración importante durante el cálculo de los factores de expansión es el estatus de las unidades seleccionadas, es decir, su clasificación en no elegibles y elegibles, estas últimas a su vez en participantes y no participantes. En la mayoría de las ocasiones, como lo fue en este caso, después de la aplicación es posible asignar dicho estatus, junto con ello definimos:

- IN_h el conjunto de UPM en la muestra del estrato h que resultaron ser no elegibles.
- Del conjunto de unidades elegibles: $S^h \backslash IN_h$,

  - Hay $m_{h_A}^*$ unidades primarias de certeza en el estrato h. De donde,
    - R_{h_A} es el conjunto con r_{h_A} UPM del estrato h que participaron en la prueba.
    - NR_{h_A} el conjunto de unidades en el estrato h que no participaron.
  - Para las UPM de no certeza,
    - R_h es el conjunto de UPM del estrato h que participaron en la prueba.
    - NR_h el conjunto de unidades en el estrato h que no participan en la prueba.

Una vez que los pesos han sido ajustados por estrato con base en la pérdida de unidades en las UPM elegibles y participantes se notó que alguno(s) de sus turno(s) no contribuyó a la prueba, para ello, y debido a que hubo censo, consideramos la selección de cada turno con probabilidad uno y se procedió a hacer ajuste por pérdida de turnos dentro de las UPM; también se puede ver como pérdida de escuelas. De esta manera se define para cada UPM i participante en la prueba del estrato explícito h,

- INt_i el conjunto de turnos no elegibles.
- Del conjunto de turnos elegibles: ${T_i}^*$ con ${t_i}^*$ elementos, se tiene que
  - Rt_ies el conjunto con rt_i turnos que participaron en la UPM i; por lo tanto, ${t_i}^* - rt_i$ es el número de turnos elegibles que no participaron en la prueba, y forman parte del conjunto ${T_i}^* \backslash Rt_i$.

Por último, para las escuelas participantes en la prueba, se define P_{hij} al conjunto de alumnos participantes y P_{hij}^C al conjunto de alumnos en la muestra y que no participaron, para el turno j de la UPM i en el estrato h.

1.21 Pesos de las UPM

A continuación se presenta la forma de los pesos de diseño en la primera etapa de muestreo y los factores de ajuste por elegibilidad o pérdida de UPM, diferenciando las unidades de certeza y las de no certeza. Además, como ya se mencionó, los marcos no siempre son perfectos y con el fin de tener mayor control sobre los participantes en escuelas más que sobre las UPM se presenta el ajuste por pérdida de turnos dentro de las UPM.

1.21.1 Pesos de diseño

Con la notación de la sección anterior, para cada UPM $i = 1, \dots, m_h$ y cada estrato explícito $h = 1, \dots, H$, el peso base de la UPM está dado por:

$$WGTBAS1_i^h = \begin{cases} 1, & si \quad i \in A_h \\ \dfrac{\sum_{U_h - A_h} N_{hi}}{(m_h - m_{h_A})N_{hi}}, & si \quad i \in U_h - A_h \end{cases}.$$

1.21.2 Factor de ajuste de las UPM

Bajo el supuesto de que la falta de respuesta de las UPM sucedió por motivos no relacionados con la variable de estudio (Little & Rubin, 2002, p.12) es posible construir un factor de ajuste dentro de cada estrato (Rutkowski, von Davier, & Rutkowski, 2014, p. 136).

Para cada UPM en el estrato explícito $h = 1, \dots, H$, el factor de ajuste por no respuesta está dado por:

$$WGTFADJ1_i^h = \begin{cases} 1 & si\ i \in IN_h \\ \frac{\sum_{i \in \{R_h \cup NR_h\}} WGTBAS1_i^h}{\sum_{i \in R_h} WGTBAS1_i^h} & si\ i \in R_h \\ \frac{m_{h_A}^*}{r_{h_A}} & si\ i \in R_{h_A} \\ 0 & si\ i \in \{NR_h \cup NR_{h_A}\}. \end{cases} \quad (1)$$

1.21.3 Factor de ajuste por pérdida de turnos dentro de las UPM

En cada escuela $j = 1, \dots, t_i$ de la UPM i participante en la prueba del estrato explícito h, el factor de ajuste tiene la forma:

$$WGTFADJt_{ij}^h = \begin{cases} 1, & si\ j \in INt_i \\ \frac{t_i^*}{rt_i}, & si\ j \in Rt_i \\ 0, & si\ j \in T_i^* \backslash Rt_i. \end{cases}$$

1.22 Pesos de los alumnos

Como ya se mencionó, dentro de cada escuela que forma parte de la muestra los alumnos fueron seleccionados mediante una estratificación dentro de los grupos escolares existentes.

1.22.1 Pesos de diseño

Si el número de alumnos en el grupo escolar g, de la escuela j en la UPM i del estrato h, es n_{ijg}^h , donde se cumple $n_{ij}^h = \sum_g n_{ijg}^h$, el peso de diseño para cada alumno $l = 1, \dots, n_{ij}^h$ en su respectivo grupo g es:

$$WGTBAS2_{ijl}^h = \frac{N_{ijg}^-}{n_{ijg}^h},$$

donde N_{ijg}^- es el número de alumnos en el grupo escolar g, tal que se cumple $N_{ij}^h = \sum_g N_{ijg}^-$.

1.22.2 Factores de ajuste

Desafortunadamente existen diversas razones por las que no todos los estudiantes seleccionados para participar en la muestra lo pueden hacer, por lo que bajo el supuesto de pérdida aleatoria se construye un nuevo factor de ajuste por no respuesta de alumnos (Rutkowski, von Davier, & Rutkowski, 2014. p. 138).

Para una escuela en particular, se considera que únicamente responden n_{ij}^{h*} de los n_{ij}^h alumnos extraídos. El factor de ajuste por no respuesta para los estudiantes es el siguiente:

$$WGTFADJ2_{ijl}^{h} = \begin{cases} \dfrac{\sum_{l \in P_{hij} \cup P_{hij}^{C}} WGTBAS2_{ijl}^{h}}{\sum_{l \in P_{hij}} WGTBAS2_{ijl}^{h}}, & si\ l \in P_{hij} \\ 0 \quad , & si\ l \in\ P_{hij}^{C}. \end{cases}$$

Una vez que los pesos han sido ajustados con base en las unidades que responden en cada una de las etapas, existen varios métodos que indican qué pesos se consideran grandes, comparados con pesos del mismo estrato, y deben acotarse. Para PLANEA EMS 2017 se acotaron aquellos que excedían 4 veces la mediana dentro de cada estrato h (como lo hace PISA en sus reportes), de tal manera que el factor de *trimming* para el alumno $l = 1, \dots, n_{ij}^{h}$ tiene la forma:

$$WGTFTRM2_{ijl}^{h} = \begin{cases} \dfrac{\sum_{l \in P^{h}} W_{l}^{h} - s_2 w_{o}^{h}}{\sum_{l \in P_{+}^{h}} W_{l}^{h}} & si\ l \in P_{+}^{h} \\ \dfrac{w_{o}^{h}}{W_{l}^{h}} & si\ l \in P^{h} \setminus P_{+}^{h}, \end{cases}$$

donde:

- $W_{l}^{h} = WGTBAS1_{i}^{h} \times WGTFADJ1_{i}^{h} \times WGTFADJt_{ij}^{h} \times WGTBAS2_{ijl}^{h} \times WGTFADJ2_{ijl}^{h}$ es el peso del alumno l en el estrato h; antes de hacer trimming.
- $w_{o}^{h} = 4(\text{Mediana de los } W_{l}^{h}\ con\ l \in P^{h})$.
- $P^{h} = \bigcup_{ijl} P_{hij}^{h}$ es el conjunto de alumnos que participaron y respondieron en el estrato h.
- P_{+}^{h} es el conjunto de UUM pertenecientes al estrato h que responden y satisfacen $W_{l}^{h} < w_{o}^{h}$.
- s_2 es el número de alumnos en el estrato h que responden y satisfacen $W_{l}^{h} \geq w_{o}^{h}$.

1.22.3 Peso final

El peso final del alumno es simplemente el producto de los pesos base junto con sus factores de ajuste por falta de respuesta y por recorte de pesos, y queda como

$$W_FSTR0_{hijl} = WGTBAS1_{i}^{h} \times WGTFADJ1_{i}^{h} \times WGTFADJt_{ij}^{h} \times WGTBAS2_{ijl}^{h} \times WGTFADJ2_{ijl}^{h} \times WGTFTRM2_{ijl}^{h},$$

para todo alumno $l = 1, \dots, n_{ij}^{h}$, de la escuela j en la UPM i del estrato h.

1.23 Criterios para publicar resultados

1.23.1 Tasa de participación

En la aplicación de PLANEA 2015 se adoptó como aceptable una tasa de participación de alumnos de 85% (criterio que utilizan estudios internacionales como PIRLS, TIMSS, etc.). Con la intención de dar resultados en la mayor cantidad posible de dominios de estudio, para la aplicación de PLANEA Educación Media Superior 2017, se propuso adoptar el criterio del Tercer Estudio Regional Comparativo y Explicativo

(TERCE), el cual es un poco más relajado. En el reporte técnico[1] del estudio se establecen tres criterios para las muestras recolectadas que deben ser cumplidos:

I. El tamaño de muestra efectiva de escuelas cubre por lo menos 80% de las escuelas esperadas en cada grado y área.
II. El tamaño de muestra efectiva incluye al menos 70% de escuelas titulares.
III. El tamaño de muestra efectiva alcanza por lo menos 80% de los alumnos que concurren efectivamente a clases.

La misma fuente también indica: "En los casos en que la población supere cualquiera de los tres criterios expuestos más arriba, se genera una nota en los reportes técnicos de tal manera de advertir en los análisis posibles sesgos que esto pudiera ocasionar". Asimismo, PIRLS y TIMSS establecen en tres categorías la situación de cada población de acuerdo con las tasas de participación, y añaden una marca a aquéllas que no cumplen con los estándares establecidos.

En PISA, cuando una población no cumple con la tasa de participación, se realizan estudios secundarios para tener elementos de decisión sobre si se reportan sus resultados o no, puesto que la falta de participación pudiese estar introduciendo sesgo en los resultados; para realizar dichos estudios se requiere tiempo (con el cual no se cuenta por ahora). Como ejemplo, en PISA 2000 en Bélgica y Holanda no se alcanzó la tasa de participación, por lo que realizaron estudios secundarios para detectar si la falta de participación introduciría sesgo en los resultados o no. Los resultados de Bélgica se publicaron porque no se encontraron motivos de preocupación respecto al potencial sesgo en los resultados por la no participación y que la tasa de participación era muy cercana a la mínima solicitada, en el caso de Holanda la tasa de participación era muy baja y los estudios secundarios realizados no detectaron si los resultados estaban sesgados o no por lo que no fueron publicados en tablas y gráficos de los reportes internacionales.

1.23.2 Precisión

La precisión en términos del valor de la razón del Error estándar del estimador de la media (EE) entre la Desviación estándar poblacional (SD): EE/SD debe ser menor o igual a 15%.

En PLANEA se recomienda publicar solamente los resultados correspondientes a los dominios que cumplen con los criterios establecidos. En el Anexo III se muestran las tasas de participación y el valor de la razón EE/SD.

1.24 Actualización sobre dominios de estudio

(Actualizado al 9 de noviembre de 2017)

El día 13 de septiembre de 2017 se solicitó valorar la pertinencia de incorporar o fusionar dominios de estudio, así como renombrarlos para la publicación de resultados. Como resultado de lo anterior, el siguiente cuadro muestra los cambios entre los dominios de estudio planeados (Nombre planeado) y los dominios de estudio en los que se publicarán resultados (Nombre final).

[1] *UNESCO-OREALC. (2016). Reporte Técnico. Tercer Estudio Regional Comparativo y Explicativo, TERCE. Santiago, Chile.*

Dominios de estudio	
Nombre planeado	Nombre final
Nacional	**Nacional**
Entidad	**Entidad**
Sostenimiento	**Tipo de control administrativo**
Autónomo	**Autónomo**
Federal	**Federal**
Estatal	**Estatal**
Privado	
Subsistemas	**Tipo de servicio**
Bachillerato estatal DGE-CGE	**Bachillerato estatal**
Bachillerato autónomo	**Bachillerato autónomo**
CECYTE	**CECYTE**
Colegio de Bachilleres	Colegio de Bachilleres
Colegio de Bachilleres México*	
CONALEP estatal	**CONALEP**
CONALEP DF-Oax.*	
DGETI	**DGETI**
DGETA	**DGETA**
Telebachilleratos	**Telebachillerato**
Telebachillerato comunitario*	**Telebachillerato comunitario**
Particulares	
	Tipo de sostenimiento
	Público
	Privado
Entidad-Sostenimiento	**Entidad-Tipo de control administrativo**
	Entidad-Tipo de sostenimiento

*Son estratos muestrales que no fueron planeados como dominios de estudio.

Apéndice A1

Este apéndice presenta las tablas correspondientes a la planeación de los tamaños muestrales según el primer esquema muestral, con los márgenes de error aproximados resumidos en la sección 12. La siguiente información no considera el cambio en el diseño muestral, ni la actualización del marco de muestreo.

Es importante tener en cuenta que (sección 8) es la razón del Error estándar del estimador de la media (EE) entre la Desviación estándar poblacional (SD): EE/SD.

Entidades

Entidad	Matrícula por entidad	EE/SD
Aguascalientes	15 952	0.0602
Baja California	39 174	0.0585
Baja California Sur	8154	0.0632
Campeche	8 944	0.0598
Chiapas	61 527	0.0534
Chihuahua	36 828	0.0588

Ciudad de México	107 679	0.0500
Coahuila	31 691	0.0564
Colima	7 727	0.0573
Durango	19 172	0.0565
Guanajuato	56 570	0.0546
Guerrero	40 600	0.0558
Hidalgo	34 581	0.0530
Jalisco	88 534	0.0547
México	166 885	0.0497
Michoacán	45 921	0.0541
Morelos	19 937	0.0566
Nayarit	11 241	0.0578
Nuevo león	51 621	0.0529
Oaxaca	40 831	0.0538
Puebla	80 317	0.0497
Querétaro	23 770	0.0557
Quintana Roo	15 134	0.0567
San Luis Potosí	32 891	0.0582
Sinaloa	34 429	0.0529
Sonora	33 139	0.0583
Tabasco	29 738	0.0566
Tamaulipas	36 639	0.0588
Tlaxcala	14 639	0.0566
Veracruz	80 610	0.0587
Yucatán	23 072	0.0581
Zacatecas	17 206	0.0535

Sostenimiento

Sostenimiento	Matrícula del sostenimiento	EE/SD
Autónomo	176 151	0.0260
Estatal	630 777	0.0167
Federal	274 648	0.0194
Particular	233 577	0.0198

Subsistemas nacionales

Subsistema	Matrícula del subsistema	EE/SD
Particulares	211 394	0.0205

COBACH	190 385	0.0285
DGETI	167 137	0.0257
Bachillerato estatal DGE-CGE	155 414	0.0496
Bachillerato Autónomo	149 654	0.0267
CECYTE	96 850	0.0379
Conalep Estatal	62 159	0.0507
Telebachillerato	51 041	0.0681

Entidad-sostenimiento

Entidad	Sostenimiento	Matrícula	EE/SD
Aguascalientes	Autónomo	1 248	0.2218
Aguascalientes	Estatal	5 925	0.1083
Aguascalientes	Federal	5 261	0.1076
Aguascalientes	Particular	3 518	0.1105
Baja California	Estatal	21 084	0.0918
Baja California	Federal	11 110	0.1056
Baja California	Particular	6 980	0.1091
Baja California Sur	Estatal	4 611	0.1015
Baja California Sur	Federal	2 750	0.1081
Baja California Sur	Particular	793	0.1198
Campeche	Autónomo	1 410	0.1806
Campeche	Estatal	4 688	0.0960
Campeche	Federal	1 921	0.1229
Campeche	Particular	925	0.1115
Chiapas	Estatal	50 859	0.0732
Chiapas	Federal	6 159	0.1111
Chiapas	Particular	4 509	0.1087
Chihuahua	Estatal	20 572	0.0918
Chihuahua	Federal	8 308	0.1106
Chihuahua	Particular	7 948	0.1112
Ciudad de México	Autónomo	22 553	0.1153
Ciudad de México	Estatal	5 247	0.1146
Ciudad de México	Federal	60 715	0.0770
Ciudad de México	Particular	19 164	0.1109
Coahuila	Autónomo	3 182	0.1237
Coahuila	Estatal	9 765	0.1135
Coahuila	Federal	7 565	0.1092
Coahuila	Particular	11 179	0.1087
Colima	Autónomo	3 816	0.0972

Entidad	Sostenimiento	Matrícula	EE/SD
Colima	Estatal	1 226	0.1161
Colima	Federal	885	0.1419
Colima	Particular	1 800	0.1155
Durango	Autónomo	927	0.1861
Durango	Estatal	9 271	0.1005
Durango	Federal	6 928	0.0998
Durango	Particular	2 046	0.1082
Guanajuato	Autónomo	2 673	0.1276
Guanajuato	Estatal	27 410	0.0935
Guanajuato	Federal	10 703	0.1114
Guanajuato	Particular	15 784	0.1115
Guerrero	Autónomo	14 808	0.1063
Guerrero	Estatal	15 603	0.1095
Guerrero	Federal	7 663	0.1190
Guerrero	Particular	2 526	0.1093
Hidalgo	Autónomo	2 782	0.1189
Hidalgo	Estatal	18 434	0.0899
Hidalgo	Federal	8 282	0.1113
Hidalgo	Particular	5 083	0.1131
Jalisco	Autónomo	51 872	0.0850
Jalisco	Estatal	15 472	0.1278
Jalisco	Federal	7 806	0.1235
Jalisco	Particular	13 384	0.1184
México	Autónomo	10 102	0.1150
México	Estatal	115 237	0.0811
México	Federal	16 162	0.1028
México	Particular	25 384	0.1089
Michoacán	Autónomo	2 040	0.1272
Michoacán	Estatal	24 073	0.0922
Michoacán	Federal	8 126	0.1085
Michoacán	Particular	11 682	0.1143
Morelos	Autónomo	1 851	0.1228
Morelos	Estatal	6 031	0.1125
Morelos	Federal	6 396	0.1077
Morelos	Particular	5 659	0.1124
Nayarit	Autónomo	3 163	0.1104
Nayarit	Estatal	3 624	0.1127
Nayarit	Federal	1 969	0.1396
Nayarit	Particular	2 485	0.1079

Entidad	Sostenimiento	Matrícula	EE/SD
Nuevo León	Autónomo	24 134	0.0934
Nuevo León	Estatal	9 177	0.1158
Nuevo León	Federal	3 286	0.1084
Nuevo León	Particular	15 024	0.1108
Oaxaca	Autónomo	1 511	0.1349
Oaxaca	Estatal	24 967	0.0852
Oaxaca	Federal	10 903	0.1121
Oaxaca	Particular	3 450	0.1119
Puebla	Autónomo	5 244	0.1117
Puebla	Estatal	53 472	0.0807
Puebla	Federal	5 798	0.1052
Puebla	Particular	15 803	0.1095
Querétaro	Autónomo	1 784	0.1409
Querétaro	Estatal	13 322	0.0869
Querétaro	Federal	2 748	0.1277
Querétaro	Particular	5 916	0.1110
Quintana Roo	Estatal	10 065	0.0794
Quintana Roo	Federal	2 837	0.1147
Quintana Roo	Particular	2 232	0.1120
San Luis Potosí	Autónomo	407	0.3123
San Luis Potosí	Estatal	13 559	0.1039
San Luis Potosí	Federal	5 882	0.0997
San Luis Potosí	Particular	13 043	0.1033
Sinaloa	Autónomo	15 266	0.0963
Sinaloa	Estatal	10 761	0.1100
Sinaloa	Federal	4 622	0.1081
Sinaloa	Particular	3 780	0.1092
Sonora	Estatal	17 847	0.0899
Sonora	Federal	10 029	0.1080
Sonora	Particular	5 263	0.1082
Tabasco	Estatal	22 638	0.0790
Tabasco	Federal	4 668	0.1153
Tabasco	Particular	2 432	0.1107
Tamaulipas	Autónomo	619	0.1959
Tamaulipas	Estatal	10 003	0.1135
Tamaulipas	Federal	16 474	0.0977
Tamaulipas	Particular	9 543	0.1113
Tlaxcala	Estatal	9 451	0.0810
Tlaxcala	Federal	3 583	0.1113

Entidad	Sostenimiento	Matrícula	EE/SD
Tlaxcala	Particular	1 605	0.1090
Veracruz	Estatal	52 658	0.0844
Veracruz	Federal	18 576	0.1160
Veracruz	Particular	9 376	0.1135
Yucatán	Autónomo	2 173	0.2226
Yucatán	Estatal	13 596	0.0901
Yucatán	Federal	3 085	0.1165
Yucatán	Particular	4 218	0.1115
Zacatecas	Autónomo	2 586	0.1298
Zacatecas	Estatal	10 129	0.0875
Zacatecas	Federal	3 448	0.1084
Zacatecas	Particular	1 043	0.1120

Nacional

Matrícula nacional	EE/SD
1 315 153	0.0099

Apéndice A2

A continuación se presentan las tablas correspondientes a la planeación de los tamaños muestrales según el esquema muestral modificado. Ver detalles en la sección 19.

Entidades

Entidad	Matrícula por entidad	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Aguascalientes	15 952	3 830	44	394	0.0498
Baja California	39 174	4 818	51	398	0.0499
Baja California Sur	8 154	3 142	41	308	0.0559
Campeche	8 944	2 856	46	301	0.0567
Chiapas	61 527	4 166	56	457	0.0466
Chihuahua	36 828	4 563	54	571	0.0415
Ciudad de México	107 679	5 776	60	440	0.0476
Coahuila	31 691	4 623	59	554	0.0421
Colima	7 727	3 084	55	394	0.0491
Durango	19 172	3 871	54	435	0.0474
Guanajuato	56 570	4 855	64	569	0.0417
Guerrero	40 565	5 341	73	615	0.0400
Hidalgo	34 581	4 886	59	548	0.0424

Jalisco	88 534	7 858	74	667	0.0386
México	166 885	7 176	78	688	0.0380
Michoacán	45 921	4 405	57	503	0.0443
Morelos	19 937	4 432	53	422	0.0482
Nayarit	11 241	3 712	55	487	0.0443
Nuevo León	51 621	6 053	68	536	0.0430
Oaxaca	40 831	4 053	67	514	0.0438
Puebla	80 315	5 783	82	823	0.0347
Querétaro	23 770	4 561	52	414	0.0487
Quintana Roo	15 134	3 472	44	352	0.0527
San Luis Potosí	32 891	3 458	40	388	0.0505
Sinaloa	34 429	6 230	70	599	0.0405
Sonora	33 139	4 337	51	351	0.0531
Tabasco	29 738	4 704	54	409	0.0491
Tamaulipas	36 639	4 929	52	500	0.0444
Tlaxcala	14 639	3 393	47	303	0.0569
Veracruz	80 610	4 231	66	603	0.0406
Yucatán	23 072	3 779	55	459	0.0462
Zacatecas	17 206	3 662	61	441	0.0470

Sostenimiento

Sostenimiento	Matrícula del sostenimiento	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Autónomo	176 151	27 450	271	1 611	0.0248
Estatal	630 777	48 907	631	5604	0.0133
Federal	274 648	41 393	398	2 259	0.0210
Particular	233 540	28 289	542	4 040	0.0156

Subsistemas Nacionales

Subsistema	Matrícula del subsistema	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Particulares	211 357	24 349	479	3 531	0.0167
COBACH	190 385	18 786	174	1 194	0.0288
DGETI	167 137	27 623	217	1 146	0.0294
Bachillerato estatal DGE-CGE	155 414	5 320	83	568	0.0419
Bachillerato Autónomo	149 654	25 694	257	1 526	0.0255
CECYTE	96 850	9 567	97	610	0.0404
Conalep EDOS	62 159	7 758	70	376	0.0514
Telebachilleratos	51 041	1 171	36	216	0.0679

Entidad-sostenimiento

Entidad	Sostenimiento	Matrícula	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Aguascalientes	Autónomo	1 248	304	2	10	0.315
Aguascalientes	Estatal	5 925	1 173	16	147	0.081
Aguascalientes	Federal	5 261	1 373	10	64	0.124
Aguascalientes	Particular	3 518	980	16	122	0.089
Baja California	Estatal	21 084	2 067	18	123	0.090
Baja California	Federal	11 110	1 740	14	87	0.107
Baja California	Particular	6 980	1 011	19	138	0.084
Baja California Sur	Estatal	4 611	1 680	16	155	0.079
Baja California Sur	Federal	2 750	992	10	55	0.133
Baja California Sur	Particular	793	470	15	86	0.102
Campeche	Autónomo	1 410	418	4	19	0.228
Campeche	Estatal	4 688	1 312	18	127	0.088
Campeche	Federal	1 921	635	7	44	0.149

Entidad	Sostenimiento	Matrícula	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Campeche	Particular	925	491	17	97	0.096
Chiapas	Estatal	50 859	2 162	27	239	0.065
Chiapas	Federal	6 159	1 026	11	56	0.133
Chiapas	Particular	4 509	978	18	138	0.084
Chihuahua	Estatal	20 572	2 264	22	288	0.059
Chihuahua	Federal	8 308	1 389	15	82	0.110
Chihuahua	Particular	7 948	910	17	137	0.085
Ciudad de México	Autónomo	22 553	1 680	13	63	0.126
Ciudad de México	Estatal	5247	1 140	15	71	0.118
Ciudad de México	Federal	60 715	1 785	15	64	0.125
Ciudad de México	Particular	19 164	1 171	17	152	0.081
Coahuila	Autónomo	3 182	922	13	68	0.120
Coahuila	Estatal	9 765	956	13	142	0.083
Coahuila	Federal	7 565	1 504	13	83	0.109
Coahuila	Particular	11 179	1 241	20	156	0.080
Colima	Autónomo	3 816	1 277	21	116	0.091
Colima	Estatal	1 226	542	12	95	0.099
Colima	Federal	885	566	8	38	0.159
Colima	Particular	1 800	699	14	111	0.092
Durango	Autónomo	927	352	6	28	0.186
Durango	Estatal	9 271	1 358	17	171	0.076
Durango	Federal	6 928	1 330	13	78	0.113
Durango	Particular	2 046	831	18	125	0.087
Guanajuato	Autónomo	2 673	842	12	57	0.131
Guanajuato	Estatal	27 410	1 140	16	140	0.084
Guanajuato	Federal	10 703	1 752	18	90	0.105
Guanajuato	Particular	15 784	1 121	18	147	0.082
Guerrero	Autónomo	14 808	2 065	18	130	0.087
Guerrero	Estatal	15 603	1 270	24	185	0.073
Guerrero	Federal	7 663	1 336	15	110	0.095
Guerrero	Particular	2 491	670	16	126	0.087
Hidalgo	Autónomo	2 782	1 078	10	48	0.143
Hidalgo	Estatal	18 434	1 372	20	158	0.079
Hidalgo	Federal	8 282	1 676	14	77	0.113
Hidalgo	Particular	5 083	760	15	126	0.088
Jalisco	Autónomo	51 872	3 964	28	218	0.068
Jalisco	Estatal	15 472	1 247	14	120	0.091

Entidad	Sostenimiento	Matrícula	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
Jalisco	Federal	7 806	1 564	17	88	0.106
Jalisco	Particular	13 384	1 083	15	144	0.083
México	Autónomo	10 102	1 524	12	58	0.131
México	Estatal	115 237	2 445	31	218	0.068
México	Federal	16 162	1 920	14	77	0.114
México	Particular	25 384	1 287	21	182	0.074
Michoacán	Autónomo	2 040	882	7	34	0.170
Michoacán	Estatal	24 073	1 256	20	157	0.080
Michoacán	Federal	8 126	1 180	13	77	0.113
Michoacán	Particular	11 682	1 087	17	132	0.087
Morelos	Autónomo	1 851	916	11	53	0.135
Morelos	Estatal	6 031	1 370	16	91	0.104
Morelos	Federal	6 396	1 184	10	61	0.127
Morelos	Particular	5 659	962	16	139	0.084
Nayarit	Autónomo	3 163	1 274	15	72	0.117
Nayarit	Estatal	3 624	846	15	130	0.086
Nayarit	Federal	1 969	604	7	86	0.105
Nayarit	Particular	2 485	988	18	147	0.080
Nuevo León	Autónomo	24 134	2 486	23	123	0.090
Nuevo León	Estatal	9 177	1 211	14	84	0.109
Nuevo León	Federal	3 286	1 215	13	79	0.111
Nuevo León	Particular	15 024	1 141	18	134	0.086
Oaxaca	Autónomo	1511	702	11	56	0.131
Oaxaca	Estatal	24 967	1 376	26	189	0.072
Oaxaca	Federal	10 903	1 376	15	96	0.102
Oaxaca	Particular	3450	599	15	98	0.100
Puebla	Autónomo	5 244	1 706	16	88	0.106
Puebla	Estatal	53 472	1 723	34	254	0.063
Puebla	Federal	5 798	1 308	12	73	0.116
Puebla	Particular	15 801	1 046	20	166	0.077
Querétaro	Autónomo	1 784	663	7	34	0.170
Querétaro	Estatal	13 322	2 135	21	164	0.078
Querétaro	Federal	2 748	806	7	36	0.166
Querétaro	Particular	5 916	957	17	125	0.088
Quintana Roo	Estatal	10 065	1 617	16	143	0.083
Quintana Roo	Federal	2 837	1 072	11	59	0.129
Quintana Roo	Particular	2 232	783	17	132	0.084

Entidad	Sostenimiento	Matrícula	Número de alumnos en la muestra	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
San Luis Potosí	Autónomo	407	152	1	5	0.444
San Luis Potosí	Estatal	13 559	1 771	18	134	0.086
San Luis Potosí	Federal	5 882	1 030	11	57	0.132
San Luis Potosí	Particular	13 043	505	10	78	0.113
Sinaloa	Autónomo	15 266	2 660	22	189	0.072
Sinaloa	Estatal	10 761	1 243	17	127	0.088
Sinaloa	Federal	4 622	1 426	13	83	0.109
Sinaloa	Particular	3 780	901	18	150	0.080
Sonora	Estatal	17 847	1 725	17	114	0.093
Sonora	Federal	10 029	1 593	14	76	0.114
Sonora	Particular	5 263	1 019	20	141	0.083
Tabasco	Estatal	22 638	2 879	28	233	0.065
Tabasco	Federal	4 668	1 063	10	66	0.122
Tabasco	Particular	2 432	762	16	116	0.091
Tamaulipas	Autónomo	619	342	4	19	0.226
Tamaulipas	Estatal	10 003	1 440	15	172	0.076
Tamaulipas	Federal	16 474	2 184	18	117	0.092
Tamaulipas	Particular	9 543	963	15	123	0.090
Tlaxcala	Estatal	9 451	1 739	19	127	0.088
Tlaxcala	Federal	3 583	1 088	11	54	0.135
Tlaxcala	Particular	1 605	566	17	104	0.095
Veracruz	Estatal	52 658	1 557	32	238	0.065
Veracruz	Federal	18 576	1 851	16	95	0.102
Veracruz	Particular	9 376	823	18	139	0.084
Yucatán	Autónomo	2 173	304	3	14	0.266
Yucatán	Estatal	13 596	1 619	23	195	0.071
Yucatán	Federal	3 085	951	12	91	0.103
Yucatán	Particular	4 218	905	17	131	0.086
Zacatecas	Autónomo	2 586	937	12	61	0.127
Zacatecas	Estatal	10 129	1 272	21	166	0.077
Zacatecas	Federal	3 448	874	11	59	0.129
Zacatecas	Particular	1 043	579	17	111	0.090

Nacional

Matrícula nacional	Número de alumnos en la muestra	Número de alumnos destinados a la prueba de "Comparabilidad"	Número de centros de trabajo en la muestra	Tamaño de muestra efectivo aproximado	EE/SD
1 315 116	146 039	7 250	1 842	15 962	0.008

Apéndice A3

(Actualizado al 9 de noviembre de 2017)

Tablas que muestran el porcentaje de participación y el valor del cociente EE/SD (precisión).

Los dominios sombreados en verde tienen una tasa de participación de al menos 80% y el valor del cociente EE/SD < 15%; los dominios con tasas de participación menores a 80% están resaltadas en rojo; los dominios con el cociente EE/SD > 0.15 van resaltados en negritas. Se recomienda no publicar los resultados de los dominios de estudio donde no se cumplen los criterios mencionados en la sección 24.

Tabla por Entidad, Entidad-Tipo de sostenimiento Entidad-Tipo de control administrativo

Entidad	Dominio de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final Ly C	Precisión final Mat
Aguascalientes	ENTIDAD		63	63	100	3830	3263	85	0	4.78	5.27
	Tipo de sostenimiento	Público	44	44	100	2850	2541	89	0	5.56	6
		Privado	19	19	100	980	722	74	1	8.45	8.89
	Tipo de control administrativo	Autónomo	4	4	100	304	285	94	0	7.77	6.64
		Estatal	21	21	100	1173	1062	91	0	5.66	6.93
		Federal	19	19	100	1373	1194	87	0	7.76	8.4
Baja California	ENTIDAD		74	74	100	4742	4057	86	0	8.54	7.57
	Tipo de sostenimiento	Público	53	53	100	3731	3204	86	0	10.76	9.78
		Privado	21	21	100	1011	853	84	0	12.45	10.58
	Tipo de control administrativo	Estatal	30	30	100	2067	1755	85	0	11.5	8.89
		Federal	**23**	**23**	**100**	**1664**	**1449**	**87**	**0**	**19.07**	**18.31**
Baja California Sur	ENTIDAD		55	55	100	3104	2652	85	0	3.7	4.88
	Tipo de sostenimiento	Público	39	39	100	2634	2236	85	0	3.65	5.07
		Privado	16	16	100	470	416	89	0	8.33	9.58
	Tipo de control administrativo	Estatal	25	25	100	1642	1465	89	0	3.89	4.61
		Federal	14	14	100	992	771	78	1	6.54	10.5
Campeche	ENTIDAD		57	57	100	2856	2472	87	0	4.9	5.24
	Tipo de sostenimiento	Público	40	40	100	2365	2094	89	0	5.28	5.59
		Privado	17	17	100	491	378	77	1	11.21	13.2
	Tipo de control administrativo	Autónomo	6	6	100	418	384	92	0	6.89	7.2
		Estatal	25	25	100	1312	1191	91	0	8.97	8.58
		Federal	9	9	100	635	519	82	0	9.34	12.41
Coahuila	ENTIDAD		77	76	99	4509	3817	85	0	8.56	8.71
	Tipo de sostenimiento	Público	49	48	98	3268	2829	87	0	7.15	5.74
		Privado	**28**	**28**	**100**	**1241**	**988**	**80**	**0**	**17.17**	**17.74**
	Tipo de control administrativo	Autónomo	13	13	100	922	735	80	0	6.02	6.53
		Estatal	15	15	100	918	866	94	0	9.58	9.17
		Federal	21	20	95	1428	1228	86	0	6.58	3.88

Entidad	Dominio de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final Ly C	Precisión final Mat
Colima	ENTIDAD		60	58	97	2945	2377	81	0	5.62	6.68
	Tipo de sostenimiento	Público	44	44	100	2374	2137	90	0	4.26	4.79
		Privado	16	14	88	571	240	42	1	18.61	23.56
	Tipo de control administrativo	Autónomo	21	21	100	1277	1164	91	0	6.64	7.08
		Estatal	13	13	100	531	474	89	0	6.72	7.65
		Federal	10	10	100	566	499	88	0	4.87	5.08
Chiapas	**ENTIDAD**		**69**	**62**	**90**	**3572**	**2918**	**82**	**0**	**18.8**	**15.12**
	Tipo de sostenimiento	**Público**	**49**	**42**	**86**	**2700**	**2243**	**83**	**0**	**19.47**	**15.72**
		Privado	20	20	100	872	675	77	1	15.84	18.62
	Tipo de control administrativo	**Estatal**	**34**	**29**	**85**	**1830**	**1568**	**86**	**0**	**20.19**	**16.21**
		Federal	15	13	87	870	675	78	1	20.55	19.37
Chihuahua	ENTIDAD		75	75	100	4563	3960	87	0	8.14	7.57
	Tipo de sostenimiento	Público	55	55	100	3653	3253	89	0	8.48	7.13
		Privado	20	20	100	910	707	78	1	22.44	23.08
	Tipo de control administrativo	Estatal	35	35	100	2264	1973	87	0	11.1	8.93
		Federal	20	20	100	1389	1280	92	0	7.81	10.11
Ciudad de México	ENTIDAD		42	40	95	2735	1871	68	1	13.85	15.06
	Tipo de sostenimiento	Público	23	23	100	1680	1229	73	1	21.47	22.26
		Privado	19	17	89	1055	642	61	1	21.4	22.53
	Tipo de control administrativo	Federal	23	23	100	1680	1229	73	1	21.47	22.26
Durango	ENTIDAD		68	68	100	3647	3188	87	0	5.89	5.17
	Tipo de sostenimiento	Público	48	48	100	2854	2482	87	0	6.48	5.61
		Privado	20	20	100	793	706	89	0	12.7	13.88
	Tipo de control administrativo	Autónomo	6	6	100	318	226	71	1	8.17	8.03
		Estatal	24	24	100	1206	1044	87	0	9.08	7.5
		Federal	18	18	100	1330	1212	91	0	9.69	9.93
Guanajuato	ENTIDAD		79	79	100	4855	4264	88	0	6.56	6.7
	Tipo de sostenimiento	Público	56	56	100	3734	3297	88	0	7.6	8.46
		Privado	23	23	100	1121	967	86	0	11.91	11.27
	Tipo de control administrativo	Autónomo	12	12	100	842	711	84	0	5.09	4.37
		Estatal	20	20	100	1140	1043	91	0	10.24	12.67
		Federal	24	24	100	1752	1543	88	0	5.9	5.69
Guerrero	ENTIDAD		92	92	100	5341	4034	76	1	4.94	5.53
	Tipo de sostenimiento	Público	76	76	100	4671	3558	76	1	5.23	5.87
		Privado	16	16	100	670	476	71	1	15.9	13.78
	Tipo de control administrativo	Autónomo	30	30	100	2065	1427	69	1	7.58	6.75
		Estatal	25	25	100	1172	1001	85	0	9.18	11.65
		Federal	21	21	100	1434	1130	79	1	11.97	12.26
Hidalgo	ENTIDAD		77	77	100	4769	4241	89	0	5.85	5.81
	Tipo de sostenimiento	Público	63	63	100	4047	3588	89	0	6.34	6.22
		Privado	**14**	**14**	**100**	**722**	**653**	**90**	**0**	**15.27**	**15.71**
	Tipo de control administrativo	Autónomo	18	18	100	1049	910	87	0	4.34	3.96
		Estatal	23	23	100	1322	1216	92	0	9.7	9.61
		Federal	22	22	100	1676	1462	87	0	8.16	7.85
Jalisco	ENTIDAD		120	119	99	7759	6546	84	0	5.96	6.48
	Tipo de sostenimiento	Público	98	98	100	6676	5752	86	0	5.87	6.4
		Privado	22	21	95	1083	794	73	1	20.1	21.4
	Tipo de control administrativo	Autónomo	55	55	100	3964	3450	87	0	7.32	7.5
		Estatal	**21**	**21**	**100**	**1163**	**956**	**82**	**0**	12.74	**18.02**
		Federal	22	22	100	1549	1346	87	0	5.08	5.24
México	ENTIDAD		104	103	99	6944	5640	81	0	4.66	5.6
	Tipo de sostenimiento	Público	81	80	99	5657	4681	83	0	5.58	6.54
		Privado	23	23	100	1287	959	75	1	11.23	9.46
	Tipo de control administrativo	Autónomo	20	19	95	1448	1164	80	0	3.29	3.15
		Estatal	36	36	100	2289	1980	87	0	6.31	7.99
		Federal	25	25	100	1920	1537	80	0	14	5.43
Michoacán	ENTIDAD		72	70	97	4087	3051	75	1	6.79	7.55
	Tipo de sostenimiento	Público	52	50	96	3000	2318	77	1	7.33	8.09
		Privado	20	20	100	1087	733	67	1	15.85	17.09
		Autónomo	14	14	100	882	533	60	1	5.19	5.4

Entidad	Dominio de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final Ly C	Precisión final Mat
	Tipo de control administrativo	Estatal	21	20	95	1018	875	86	0	10.62	11.42
		Federal	17	16	94	1100	910	83	0	10.27	12.98
Morelos	ENTIDAD		67	64	96	4292	3779	88	0	7.41	8.23
	Tipo de sostenimiento	Público	50	49	98	3404	3017	89	0	4.95	5.27
		Privado	**17**	**15**	**88**	**888**	**762**	**86**	**0**	**20.34**	**21.83**
	Tipo de control administrativo	Autónomo	13	13	100	916	790	86	0	3.82	3.38
		Estatal	21	20	95	1304	1184	91	0	6.87	4.25
		Federal	16	16	100	1184	1043	88	0	9.26	11.34
Nayarit	ENTIDAD		70	68	97	3560	2475	70	1	5.48	4.98
	Tipo de sostenimiento	Público	49	47	96	2610	2070	79	1	5.08	3.93
		Privado	21	21	100	950	405	43	1	15.57	15.73
	Tipo de control administrativo	Autónomo	20	19	95	1198	933	78	1	3.71	4.2
		Estatal	19	19	100	846	710	84	0	13.45	10.27
		Federal	10	9	90	566	427	75	1	10.09	8.29
Nuevo León	ENTIDAD		90	89	99	5977	4822	81	0	3.96	4.99
	Tipo de sostenimiento	Público	70	69	99	4836	4165	86	0	2.64	3.01
		Privado	20	20	100	1141	657	58	1	13	16.12
	Tipo de control administrativo	Autónomo	33	32	97	2410	2072	86	0	3.19	3.74
		Estatal	19	19	100	1211	995	82	0	4.79	8.15
		Federal	18	18	100	1215	1098	90	0	3.9	4.54
Oaxaca	ENTIDAD		74	73	99	3995	3259	82	0	5.64	6.42
	Tipo de sostenimiento	Público	59	59	100	3417	2769	81	0	6.16	7.15
		Privado	15	14	93	578	490	85	0	12.59	9.86
	Tipo de control administrativo	Autónomo	13	13	100	665	429	65	1	6.2	5.42
		Estatal	27	27	100	1376	1210	88	0	8.41	9.35
		Federal	19	19	100	1376	1130	82	0	7.36	11.36
Puebla	ENTIDAD		89	89	100	5340	4839	91	0	4.34	5.21
	Tipo de sostenimiento	Público	75	75	100	4661	4250	91	0	4.66	5.53
		Privado	14	14	100	679	589	87	0	13.43	**15.7**
	Tipo de control administrativo	Autónomo	23	23	100	1630	1473	90	0	5.91	6.77
		Estatal	35	35	100	1723	1599	93	0	5.86	7
		Federal	17	17	100	1308	1178	90	0	5.81	7.81
Querétaro	ENTIDAD		72	72	100	4561	3856	85	0	5.57	5.66
	Tipo de sostenimiento	Público	54	54	100	3604	3030	84	0	7.75	7.79
		Privado	18	18	100	957	826	86	0	3.71	4.76
	Tipo de control administrativo	Autónomo	10	10	100	663	568	86	0	5.23	5.52
		Estatal	33	33	100	2135	1798	84	0	11.23	11.44
		Federal	11	11	100	806	664	82	0	6.86	8.19
Quintana Roo	ENTIDAD		60	60	100	3472	3022	87	0	4.25	4
	Tipo de sostenimiento	Público	43	43	100	2689	2358	88	0	4.76	4.33
		Privado	17	17	100	783	664	85	0	11.22	10.51
	Tipo de control administrativo	Estatal	27	27	100	1617	1440	89	0	5.86	5.3
		Federal	16	16	100	1072	918	86	0	5.04	5.96
San Luis Potosí	ENTIDAD		51	51	100	3382	3137	93	0	11.55	13.93
	Tipo de sostenimiento	Público	41	41	100	2877	2672	93	0	6.5	6.49
		Privado	**10**	**10**	**100**	**505**	**465**	**92**	**0**	**25.29**	**30.25**
	Tipo de control administrativo	Autónomo	2	2	100	152	135	89	0	8.76	8.32
		Estatal	26	26	100	1771	1646	93	0	8.22	8.7
		Federal	13	13	100	954	891	93	0	10.09	7.48
Sinaloa	ENTIDAD		104	103	99	6078	5162	85	0	3.5	3.44
	Tipo de sostenimiento	Público	85	85	100	5329	4488	84	0	3.5	3.58
		Privado	**19**	**18**	**95**	**749**	**674**	**90**	**0**	**16.95**	13.95
	Tipo de control administrativo	Autónomo	42	42	100	2660	2034	76	1	5.77	5.06
		Estatal	23	23	100	1243	1109	89	0	6.17	7.82
		Federal	20	20	100	1426	1345	94	0	5.36	6.45
Sonora	ENTIDAD		65	65	100	4092	3571	87	0	5.39	5.49
	Tipo de sostenimiento	Público	45	45	100	3073	2691	88	0	5.93	5.95
		Privado	20	20	100	1019	880	86	0	13.45	14.12
	Tipo de control administrativo	Estatal	26	26	100	1708	1558	91	0	7.09	6.69
		Federal	19	19	100	1365	1133	83	0	11.35	12.45
Tabasco	ENTIDAD		77	76	99	4666	4041	87	0	4.16	4.03

Entidad	Dominio de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final Ly C	Precisión final Mat
	Tipo de sostenimiento	Público	60	60	100	3942	3402	86	0	4.54	4.42
		Privado	17	16	94	724	639	88	0	14.04	12.08
	Tipo de control administrativo	Estatal	45	45	100	2879	2553	89	0	5.3	5.29
		Federal	15	15	100	1063	849	80	0	5.97	4.9
Tamaulipas	ENTIDAD		72	72	100	4842	4197	87	0	6.33	6.39
	Tipo de sostenimiento	Público	56	56	100	3890	3454	89	0	5.88	5.47
		Privado	16	16	100	952	743	78	1	16.34	17.33
	Tipo de control administrativo	Autónomo	5	5	100	342	251	73	1	6.26	6.74
		Estatal	21	21	100	1364	1190	87	0	11.53	10.68
		Federal	30	30	100	2184	2013	92	0	6.38	6.07
Tlaxcala	ENTIDAD		59	59	100	3363	2952	88	0	5.08	4.23
	Tipo de sostenimiento	Público	42	42	100	2797	2456	88	0	5.66	4.6
		Privado	17	17	100	566	496	88	0	12.21	11.84
	Tipo de control administrativo	Estatal	27	27	100	1721	1535	89	0	7.74	6.23
		Federal	15	15	100	1076	921	86	0	4.49	5.48
Veracruz	ENTIDAD		80	80	100	4231	3883	92	0	7.88	8.02
	Tipo de sostenimiento	Público	60	60	100	3400	3150	93	0	8.96	9.08
		Privado	20	20	100	831	733	88	0	13.89	13.73
	Tipo de control administrativo	Estatal	34	34	100	1549	1448	93	0	11.85	12.08
		Federal	26	26	100	1851	1702	92	0	6.69	8.33
Yucatán	ENTIDAD		63	63	100	3745	3174	85	0	5.54	6.17
	Tipo de sostenimiento	Público	46	46	100	2840	2377	84	0	6.11	6.98
		Privado	17	17	100	905	797	88	0	8	9.03
	Tipo de control administrativo	Autónomo	4	4	100	304	270	89	0	6.81	7.32
		Estatal	27	27	100	1585	1347	85	0	8.68	10.1
		Federal	15	15	100	951	760	80	0	6.49	6.89
Zacatecas	ENTIDAD		67	67	100	3662	3180	87	0	4.16	4.98
	Tipo de sostenimiento	Público	50	50	100	3083	2661	86	0	4.44	5.35
		Privado	17	17	100	579	519	90	0	11.47	8.93
	Tipo de control administrativo	Autónomo	16	16	100	937	746	80	0	4.4	5.18
		Estatal	22	22	100	1272	1136	89	0	6.64	8.35
		Federal	12	12	100	874	779	89	0	8.07	6.9

Tabla por Tipo de sostenimiento, Tipo de control administrativo y Tipo de servicio

Dominio de estudio		Escuelas planeadas	Escuelas participantes	Escuelas Tasa de participación (%)	Alumnos planeados en escuelas participantes	Alumnos participantes en escuelas participantes	Alumnos Tasa de participación (%)	TR<80%	Precisión final Ly C	Precisión final Mat
Tipo de sostenimiento	Público	1755	1740	99	112346	96452	86	0	3.2	2.5
	Privado	589	579	98	27170	21248	78	1	3.6	3.9
Tipo de Tipo de control administrativo	Autónomo	380	377	99	25366	20690	82	0	2.4	2.5
	Estatal	800	793	99	46351	40927	88	0	4.2	3.5
	Federal	575	570	99	40629	34835	86	0	6.1	4.5
Tipo de servicio	Bachillerato estatal	92	91	99	5020	4326	86	0	5.2	6.4
	Bachillerato autónomo	380	377	99	25366	20690	82	0	2.4	2.5
	CECYTE	**140**	**137**	**98**	**9263**	**8395**	**91**	**0**	**18.7**	**15.3**
	Colegio de Bachilleres	277	277	100	19014	16684	88	0	5.5	2.6
	CONALEP	124	124	100	8560	7339	86	0	3.6	3.2
	DGETA	89	88	99	5670	4902	86	0	5.6	5.2
	DGETI	368	365	99	27189	23624	87	0	2.3	2.7
	Telebachillerato	36	34	94	1137	1016	89	0	8.1	9.1
	Telebachillerato comunitario	52	52	100	862	793	92	0	7.5	7.1

Nota: Los dominios sombreados en verde tienen una tasa de participación de al menos 80% y el valor del cociente EE/SD<= 15%; los dominios con tasas de participación menores a 80% están resaltados en rojo; los dominios con el cociente EE/DE > 0.15 se encuentran resaltados en negritas.

Tablas resumen donde se indica en qué dominios de estudio sí se cumplen los criterios establecidos y en cuáles no.

Tabla por Entidad, Entidad-Tipo de sostenimiento Entidad-Tipo de control administrativo

Entidad federativa	Dominio de estudio*										
	Entidad	Tipo de sostenimiento				Tipo de control administrativo					
		Público		Privado		Autónomo		Estatal		Federal	
		LyC	Mat	LyC	Mat	LyC	Mat	LyC	Mat	LyC	Mat
Aguascalientes	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Baja California	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	No	No
Baja California Sur	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	No	No
Campeche	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Coahuila	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Colima	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Chiapas	No	No	No	No	No	NA	NA	No	No	No	No
Chihuahua	Sí	Sí	Sí	No	No	NA	NA	Sí	Sí	Sí	Sí
Ciudad de México	No	No	No	No	No	NP	NP	NP	NP	No	No
Durango	Sí	Sí	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí
Guanajuato	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Guerrero	No	No	No	No	No	No	No	Sí	Sí	No	No
Hidalgo	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Jalisco	Sí	Sí	Sí	No	No	Sí	Sí	Sí	No	Sí	Sí
México	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Michoacán	No	No	No	No	No	No	No	Sí	Sí	Sí	Sí
Morelos	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Nayarit	No	No	No	No	No	No	No	Sí	Sí	No	No
Nuevo León	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Oaxaca	Sí	Sí	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí
Puebla	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí
Querétaro	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Quintana Roo	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	Sí	Sí
San Luis Potosí	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Sinaloa	Sí	Sí	Sí	No	No	No	No	Sí	Sí	Sí	Sí
Sonora	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	Sí	Sí
Tabasco	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	Sí	Sí
Tamaulipas	Sí	Sí	Sí	No	No	No	No	Sí	Sí	Sí	Sí
Tlaxcala	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	Sí	Sí
Veracruz	Sí	Sí	Sí	Sí	Sí	NA	NA	Sí	Sí	Sí	Sí
Yucatán	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Zacatecas	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí

Tabla por Tipo de sostenimiento, Tipo de control administrativo y Tipo de servicio

Dominio de estudio		Resultado*	
		LyC	Mat
Tipo de sostenimiento	Público	SÍ	SÍ
	Privado	NO	NO
Tipo de control administrativo	Autónomo	SÍ	SÍ
	Estatal	SÍ	SÍ
	Federal	SÍ	SÍ
Tipo de servicio	Bachillerato estatal	SÍ	SÍ
	Bachillerato autónomo	SÍ	SÍ
	CECYTE	NO	NO
	Colegio de Bachilleres	SÍ	SÍ
	CONALEP	SÍ	SÍ
	DGETA	SÍ	SÍ
	DGETI	SÍ	SÍ
	Telebachillerato	SÍ	SÍ
	Telebachillerato comunitario	SÍ	SÍ

*Criterio basado en las tasas de participación redondeadas (al menos de 80%) y el valor del cociente EE/SD<=15%.

NA (No aplica): Se refiere a que no hay centros escolares en Tipo de control administrativo autónomo en dichas entidades.
NP (No participó): En el caso de la Cuidad de México los centros escolares en Tipo de control administrativo autónomo y estatal no participaron en la aplicación de PLANEA EMS 2017.

NO: Tiene tasa de participación menor a 80% y/o el valor del cociente EE/DS >15%.

SÍ: Tiene tasa de participación mayor a 80% y el valor del cociente EE/DS <=15%.

Anexo 2. Metodología de escalamiento para la Evaluación de Logro referida al Sistema Educativo Nacional (ELSEN)

En su modalidad de Evaluación de Logro referida al Sistema Educativo Nacional, el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA) es el programa para la evaluación del logro de los aprendizajes de los estudiantes en los distintos niveles de la educación obligatoria y forma parte del Sistema Nacional de Evaluación Educativa. Permite evaluar un conjunto amplio de contenidos del currículo y se aplica a muestras representativas del país de todos los grados terminales de la educación básica, en ciclos de cuatro años. Sus resultados sirven para monitorear al Sistema Educativo Nacional (SEN) y, en conjunto, aportan información a los tomadores de decisiones de política educativa. La aplicación de las evaluaciones del logro del SEN están a cargo del INEE.

El propósito de PLANEA SEN no es obtener estimaciones precisas de logro (habilidad) de los estudiantes en lo individual, sino de grupos o poblaciones de estudiantes. En este caso, se quiere monitorear el progreso de las poblaciones y con ese fin es necesario calcular varios estadísticos sobre la variable de logro como promedio, desviación estándar, porcentaje, percentil, etc. En este sentido, es conveniente construir las escalas de calificación mediante la metodología de los valores plausibles (Mislevy, 1991). Esta metodología está diseñada para cumplir el propósito de la prueba y es ampliamente utilizada internacionalmente en pruebas de rendimiento como PISA (*Programme for International Student Assessment*), TIMSS (*Trends in International Mathematics and Science Study*), PIRLS (*Progress in International Reading Literacy Study*), etcétera.

Mislevy (1991) señala que al utilizar métodos tradicionales de escalamiento en los que primero se obtienen estimaciones individuales de logro y con éstas se calculan estadísticas de grupo, se pueden generar resultados sesgados. Además, PLANEA SEN evalúa un conjunto amplio de contenidos del currículo mediante un diseño matricial en el que cada alumno evaluado únicamente contesta un subconjunto del total de reactivos que integran la evaluación.

El logro de los estudiantes es una variable que no se puede observar directamente, por lo que se mide mediante las respuestas a los reactivos que cada uno resuelve, dice Mislevy (1991), y advierte que los procedimientos estándar para hacer inferencias, provenientes de muestreos complejos, no aplican cuando la variable de interés no puede ser observada directamente. Esta es otra razón por la que se adopta la metodología de los valores plausibles para la construcción de las escalas de calificación de PLANEA SEN.

Cabe señalar que las puntuaciones derivadas de esta metodología tienen algunas propiedades paradójicas, puesto que no sirven como puntuaciones individuales y tampoco para el diagnóstico de los sujetos, sino que sólo se pueden utilizar para la estimación precisa de parámetros de la población.

A continuación se describe la implementación de la metodología de escalamiento con valores plausibles para PLANEA SEN.

2.1 Los valores plausibles

La metodología de escalamiento con valores plausibles se combina con el escalamiento estándar de la Teoría de respuesta al ítem, el cual supone generalmente que la habilidad es una cantidad fija. En la metodología de estimación con valores plausibles, se asume que la habilidad o logro de cualquier estudiante puede describirse mediante una distribución de probabilidad, de modo que los valores plausibles son aquellos posibles valores de habilidad extraídos aleatoriamente de su distribución empírica. Dicha distribución tiene información proveniente de los reactivos que respondió el estudiante y del cuestionario de contexto.

Usualmente es suficiente hacer cinco extracciones de cada uno de los estudiantes para obtener estimaciones fiables de los parámetros de la población. Sin embargo, en la aplicación de educación media superior se decidió aumentar a diez extracciones para reducir aún más la incertidumbre producida por el uso de estos valores.

Para describir de manera formal la metodología, es necesario introducir un poco más de notación. Supóngase que $\boldsymbol{Y}$ representa las respuestas del cuestionario de contexto de todos los estudiantes que fueron examinados en la muestra, también suponga que $\boldsymbol{X}$ representa las respuestas de los estudiantes a los reactivos para medir el logro y, finalmente, suponga que $\boldsymbol{\theta}$ representa los valores de la variable latente.

El principio con el que funcionan estos valores aleatorios para estimar características de la población es el siguiente:

Supóngase que los valores de $\boldsymbol{\theta}$ fueran observables directamente en la muestra, entonces sería posible calcular el estadístico $Q = Q(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y})$ que puede ser una media, un percentil, un coeficiente de regresión o cualquier otro estadístico que depende de las respuestas de contexto o de las respuestas de los estudiantes a los reactivos o, inclusive, de los valores de la variable latente. En tal caso, se pueden utilizar métodos estándares para hacer estadística.

Sin embargo, los valores de $\boldsymbol{\theta}$ no pueden observarse directamente en ninguna muestra. Por lo que se propone que se asuma que los valores de $\boldsymbol{\theta}$ están perdidos de manera completamente aleatoria (*Missing Completely at Random*). Este supuesto hace posible el uso de métodos de imputación múltiple para estimar Q.

Bajo esta perspectiva Rubin (1977) afirma que el estadístico $Q(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y})$ puede ser aproximado por el valor esperado del estadístico condicionado a que se observen los valores de $\boldsymbol{X}$ y $\boldsymbol{Y}$; es decir:

$$Q(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y}) \approx E(Q|\boldsymbol{X}, \boldsymbol{Y})$$

Siguiendo la definición general de valor esperado se tiene que:

$$E(Q|\boldsymbol{X}, \boldsymbol{Y}) = \int Q(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y}) p(\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{Y}) d\boldsymbol{\theta} \ldots (1)$$

En esta definición, los valores $\boldsymbol{X}$ y $\boldsymbol{Y}$ no tienen ningún problema de ser observados en cualquier muestra y se deduce que para calcular cualquier estadístico que depende de variables latentes, se necesita extraer valores aleatorios de la distribución $p(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y})$, después se necesita evaluar $Q(\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Y})$ en dichos valores y finalmente se deben promediar los estadísticos evaluados en los valores aleatorios para obtener una sola estimación del estadístico de interés.

El procedimiento de estimación con valores plausibles se puede describir con cuatro etapas, el análisis preliminar del comportamiento de los reactivos que conforman la prueba, la calibración de los reactivos, el condicionamiento con la estimación de parámetros poblacionales y, finalmente, la extracción de valores plausibles. A continuación, se describirán cada una de estas etapas.

2.2 Análisis preliminar al escalamiento

Una vez que las pruebas fueron administradas a los alumnos muestreados y sus respuestas fueron integradas a archivos electrónicos, es necesario llevar a cabo un análisis de los reactivos con la finalidad de verificar el comportamiento de cada una de las preguntas que se incluyeron en los análisis definitivos. Aunque las preguntas ya fueron puestas a prueba con anterioridad en la fase de piloteo, se repite el procedimiento debido a los ajustes realizados en las propias preguntas.

El análisis se dividió en dos partes:

- Análisis con un modelo de Teoría de respuesta al ítem
- Detección de funcionamiento diferencial de reactivos

El modelo aplicado en PLANEA SEN es una forma generalizada del modelo logístico simple de Rasch, este modelo se conoce como modelo logit multinomial con coeficientes aleatorios, que se describe en Adams, R., Wilson, M. y Wang, W. (1997). Se implementa con el software ConQuest 2 (Adams & Wu, 2002). En este modelo los reactivos se pueden describir por un conjunto de parámetros fijos desconocidos y se asume que la habilidad de los sustentantes tiene una distribución de probabilidad, lo que será descrito más adelante en el modelo de la población.

El modelo logístico simple predice la probabilidad de responder correctamente un reactivo, basado en la habilidad del sustentante i y la dificultad del reactivo δ_j. Es decir,

$$p(x_{ij}|\theta_i,\delta_j) = \frac{\exp\left[x_{ij}(\theta_i-\delta_j)\right]}{1+\exp\left(\theta_i-\delta_j\right)} \ldots (2)$$

Donde:
x_{ij} Es la respuesta del estudiante i al reactivo j, 1 si es correcta y 0 si es incorrecta;
θ_i Es el nivel de logro del estudiante i en la escala *logit*;
δ_j Es el parámetro de localización del reactivo j, caracterizando la dificultad del reactivo.
La indeterminación de los parámetros del modelo se resuelve mediante la imposición de la siguiente restricción:

$$\sum_{j=1}^{n} \delta_j = 0$$

En donde n es el total de reactivos de la prueba.

2.2.1. Análisis con el modelo logístico de Rasch

En PLANEA SEN se evaluaron dos asignaturas:

- Lenguaje y Comunicación
- Matemáticas

En el año terminal de bachillerato, para cada una de las asignaturas se ajustó el modelo logístico de Rasch de manera totalmente independiente. Una vez hecho esto, el análisis consistió en revisar principalmente tres indicadores:

- Correlación punto biserial (r_{pb})
- Ajuste al modelo de Rasch ($infit$)
- Curva característica del reactivo

Los reactivos fueron clasificados de la siguiente forma:

Según su poder de discriminación
de acuerdo con la correlación punto biserial de la respuesta correcta

Rango de discriminación	Clasificación
$r_{pb} \geq 0.15$	El reactivo discrimina de forma aceptable
$r_{pb} < 0.15$	El reactivo debe ser eliminado de la prueba

Según el ajuste del modelo logístico de Rasch

Rango de discriminación	Clasificación
$0.8 \leq infit \leq 1.2$	El reactivo ajusta de forma aceptable
$infit < 0.8$ o $1.2 < infit < 2$	El reactivo desajusta moderadamente
$infit \geq 2$	El reactivo debe ser eliminado de la prueba

Según la curva característica

Rango de discriminación	Clasificación
La curva característica formada con los datos observados está cerca de la curva característica esperada	El reactivo ajusta de forma aceptable
La curva característica formada con los datos observados no se parece de la curva característica esperada	El reactivo debe ser eliminado de la prueba

2.2.2 Detección del funcionamiento diferencial de los reactivos (DIF) mediante el modelo logístico de Rasch

El funcionamiento diferencial de reactivos (DIF) ocurre cuando diferentes grupos de estudiantes que responden un reactivo, con el mismo nivel de habilidad, no tienen la misma probabilidad de contestar correctamente; es decir, dicho reactivo funciona de forma distinta según el contexto y el grupo al que se aplique, después de mantener fijo el nivel de habilidad.

Es común que estos grupos sean determinados por variables como el género, características étnicas, culturales, geográficas, etc. Históricamente, se construyen dos grupos de comparación: el grupo de referencia, integrado por los individuos a los cuales se espera que la prueba aplicada les favorezca; y el grupo focal, que consiste en los individuos para los cuales se espera que la prueba aplicada los ponga en desventaja.

Si los datos ajustan razonablemente al modelo logístico simple de Rasch, se puede demostrar que la comparación de las probabilidades de acierto para cada reactivo en los grupos de interés, bajo la hipótesis de no DIF, es equivalente a la comparación de los parámetros de dificultad calculados independientemente en cada uno de los grupos a contrastar.

El método implementado para analizar el funcionamiento diferencial de reactivos compara dos grupos mediante la estimación de la dificultad de los reactivos para cada uno de los grupos, es decir, se estiman $\hat{\delta}_{j,F}$, $\hat{\delta}_{j,R}$, y sus respectivos errores estándar, donde se considera al grupo focal (F) y al grupo de referencia (R). (Wright, B.D., & Stone, M.H., 1979), (Wright, B.D., Mead, R.J, & Draba, R., 1976) y (Wright, B.D., & Masters, G.N., 1982).

Para la prueba PLANEA SEN 2017 de nivel bachillerato, tanto para la asignatura de Lenguaje y Comunicación como para Matemáticas, se realizó un análisis DIF en los grupos de estudio definidos por la variable sexo; el grupo focal o minoritario fue la subpoblación de mujeres y el grupo de referencia fue la subpoblación de hombres.

Es importante señalar que el tamaño de las subpoblaciones se relaciona directamente con la potencia del análisis DIF de la prueba. Por lo que, previo a éste, se verificó que el tamaño de las subpoblaciones, tanto de la focal como la de referencia, tuviera al menos 500 sustentantes (Alice Guilleux *et al*., 2014).

2.2.3 Los análisis DIF de PLANEA SEN

El método de Rasch para detectar DIF consistió en la estimación de los parámetros de dificultad de manera aislada en cada uno de los grupos de interés. La diferencia media en logro entre ambos grupos se controló fijando la dificultad promedio de los reactivos en cero:

$$\sum_{j=1}^{n} \hat{\delta}_{j,R} = \sum_{j=1}^{n} \hat{\delta}_{j,F} = 0$$

Se calculó la diferencia en la escala *logit* entre las dificultades de reactivos estimadas para ambos grupos, como una medida para determinar el funcionamiento diferencial de reactivos:

$$\hat{\Delta}_j = \hat{\delta}_{j,R} - \hat{\delta}_{j,F}$$

Se calculó el índice estandarizado que está dado por:

$$Z_j = \frac{\hat{\Delta}_j}{\sqrt{(SE_{\hat{\delta}_{j,R}})^2 + (SE_{\hat{\delta}_{j,F}})^2}}$$

Bajo la hipótesis de que no se presenta DIF, el valor esperado de $\hat{\Delta}_j$ es 0, y Z_j tiene media 0 y varianza 1. Por lo cual, se reportan solamente los reactivos que presentan una diferencia significativa, es decir, aquellos en donde: $Z_j \leq -1.96$ o $Z_j \geq 1.96$.

Una vez identificados los reactivos con funcionamiento diferencial significativo, se procede a clasificarlos en tres categorías, bajo el siguiente criterio (Wilson, 2005, p.167):

Categoría	Criterio
DIF despreciable "A"	$\lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert < 0.426$
DIF intermedio "B"	$0.426 \leq \lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert < 0.638$
DIF grande "C"	$\lvert \hat{\delta}_{j,R} - \hat{\delta}_{j,F} \rvert \geq 0.638$

En los casos en que la diferencia es significativa, es decir, se presenta DIF, el grupo focal de cada análisis se ve favorecido si $\hat{\Delta}_j$ y Z_j son valores positivos.

2.2.4 Eliminación de reactivos

Una vez clasificados los reactivos de acuerdo con los problemas que presentaron de discriminación, ajuste, forma de la curva característica o funcionamiento diferencial. La Dirección General de Medición y Tratamiento de Datos tomó la decisión sobre qué reactivos eliminar. Las cantidades se presentan a continuación:

Asignatura	Reactivos eliminados	Reactivos utilizados
Lenguaje y Comunicación	2	148
Matemáticas	2	148

2.2.5 El modelo condicional de respuesta al ítem

De los supuestos de los modelos IRT, el más importante es la independencia condicional. Bajo este supuesto, las probabilidades de los reactivos dependen solamente de θ y de su dificultad, y no son afectadas por otras variables como las características sociodemográficas. Denotando estas variables para cada uno de los estudiantes con y.

Bajo este supuesto y utilizando (2), la probabilidad conjunta de un patrón de respuesta $x' = (x_1, \dots, x_n)$, a través de los n reactivos, está dada por:

$$p(\boldsymbol{x}|\theta, \boldsymbol{\delta}, \boldsymbol{y}) = p(\boldsymbol{x}|\theta, \boldsymbol{\delta}) = \prod_{j=1}^{n} p\left(x_j|\theta, \delta_j\right) \qquad \dots (3)$$

Donde:

$\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$, es el vector de parámetros de dificultad de los n reactivos.

Los parámetros de dificultad se estiman en la calibración y, posterior a ese paso, se consideran valores fijos. La respuesta de cualquier subconjunto de reactivos induce una función de verosimilitud para θ y si, adicionalmente, bajo el supuesto de que existe independencia entre los sujetos, utilizando (3), obtenemos "el modelo condicional de respuesta al ítem" o "el modelo de la variable latente".

$$p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{Y}) = p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\delta}) = \prod_{i=1}^{N} p(\boldsymbol{x}_i|\theta_i, \boldsymbol{\delta}) \qquad \dots . (4)$$

Donde:
N es la cantidad total de estudiantes en la muestra.

2.3 Calibración de los reactivos

Una vez depuradas las claves de respuesta, se eliminaron los reactivos con grandes deficiencias. La fase de calibración de los instrumentos consistió en la estimación de los parámetros de dificultad de los reactivos para el modelo logístico simple de Rasch.

Para la calibración, el modelo condicional de respuesta al ítem es usado en conjunción del modelo de la población, pero sin utilizar variables de condicionamiento. Más adelante se habla sobre dichas variables.

El método utilizado para hacer la estimación de los parámetros de los reactivos fue el de máxima verosimilitud marginal (Bock y Aitkin, 1981), en el cual se asume independencia condicional entre las respuestas a los reactivos y la habilidad como un efecto aleatorio. De esta forma es posible eliminar los parámetros de las personas del proceso de estimación integrando sobre la habilidad, lo cual se reduce a maximizar el logaritmo de la verosimilitud que surge de:

$$p(\boldsymbol{X}|\boldsymbol{\alpha},\boldsymbol{\delta}) = \prod_{i=1}^{N} \int p(\boldsymbol{x}_i|\theta,\boldsymbol{\delta}) p(\theta|\boldsymbol{\alpha})\, d\theta$$

La calibración de cada una de las escalas se realizó independientemente de los parámetros de las otras; se hizo utilizando la máxima cantidad de información disponible, es decir, con toda la muestra de los estudiantes.

2.3.1 Análisis multidimensional de Rasch

Posterior al paso de calibración, se aprovechó que el modelo de la población hace posible ajustar un modelo bidimensional para las habilidades, en el cual se consideran relacionadas las habilidades de un estudiante en Lenguaje y Comunicación con Matemáticas. En la siguiente sección se explica el modelo de la población. Del ajuste del modelo bidimensional, se calculan puntajes basados en métodos de máxima verosimilitud y se deriva que la correlación de los puntajes entre las dimensiones fue el siguiente:

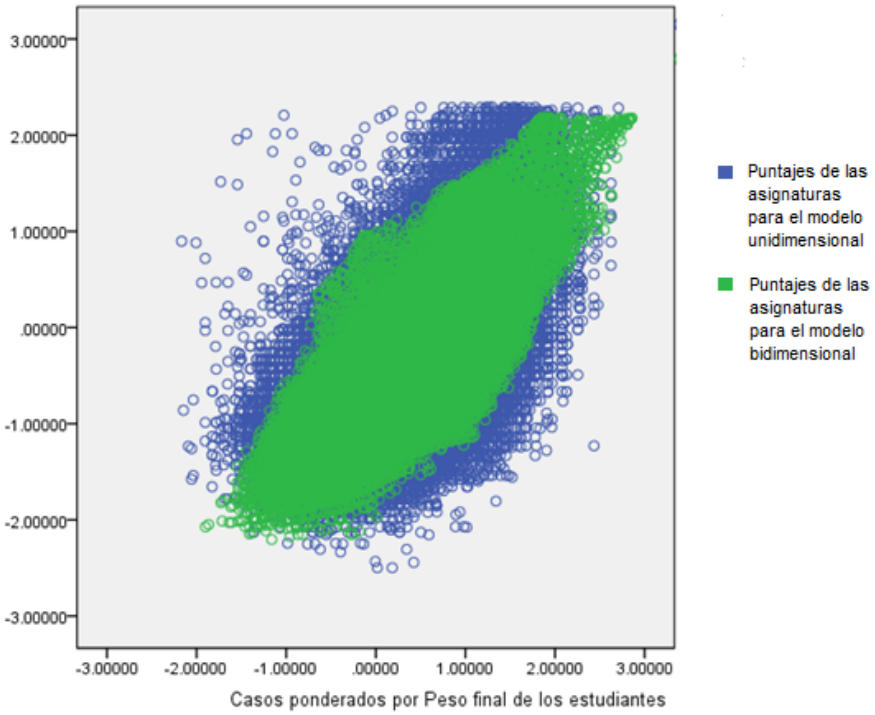
Correlación	Matemáticas
Lenguaje y Comunicación	0.725

En ambas asignaturas la confiabilidad de las pruebas aumentó.

Asignatura	Confiabilidad unidimensional sin condicionamiento	Confiabilidad bidimensional sin condicionamiento
Lenguaje y Comunicación	0.817	0.840
Matemáticas	0.897	0.898

En los cuadros anteriores se observó que los puntajes de las asignaturas tienen una considerable asociación lineal y que la confiabilidad de las pruebas se incrementa con respecto a los modelos unidimensionales. Para verificar que existe una alta correlación entre los puntajes de ambas asignaturas, se obtuvieron estimaciones de las habilidades con los modelos unidimensionales o marginales y, posteriormente, con el modelo bidimensional. La gráfica se muestra a continuación.

Gráfica de puntajes entre asignaturas



Por lo tanto, se tomó la decisión de utilizar un modelo bidimensional para generar los puntajes de ambas asignaturas.

2.3.2 El modelo de la población

Para continuar con el proceso de escalamiento, es necesario suponer que existe una distribución de probabilidad asociada a la variable latente y que tiene una forma funcional conocida.

Mislevy (1991) supone que la distribución de la variable latente θ, dadas las variables sociodemográficas $\boldsymbol{y}$, tiene una forma funcional conocida que denotaremos con $p(\theta|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha})$, que está caracterizada por un conjunto de parámetros desconocidos $\boldsymbol{\alpha}$. La práctica más común es asumir que las habilidades de los estudiantes provienen de una distribución normal multivariada.

Por su parte, Adams, Wilson, y Wu (1997) opinan que es muy natural reemplazar la media de la distribución normal multivariada con un modelo de regresión $\Gamma'\boldsymbol{y}_i$, donde $\boldsymbol{y}_i$ es un vector de valores fijos del i-ésimo estudiante.

Entonces, el modelo poblacional del estudiante i será:

$$\theta_i = \Gamma'\boldsymbol{y}_i + \boldsymbol{\epsilon}$$

Donde $\boldsymbol{\epsilon}$ se distribuye normal multivariada con media $\mathbf{0}$ y matriz de varianza y covarianza Σ entre las dimensiones. De esta forma, los parámetros a estimar están en $\boldsymbol{\alpha} = (\Gamma, \Sigma)$. En este contexto las $\boldsymbol{y}_i$ son

conocidas como **variables de condicionamiento,** que resultan indispensables para la estimación final de las habilidades.

De esta forma, el modelo de la población se define asumiendo que existe independencia entre los sustentantes:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = \prod_{i=1}^{N} p(\theta_i|\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\delta}) \quad \ldots(5)$$

En el modelo de la población se estimaron los coeficientes de regresión y la matriz de varianza y covarianza, en la fase del condicionamiento y la estimación de parámetros.

2.4 El condicionamiento y la estimación de parámetros poblacionales

Las variables de condicionamiento tienen como meta maximizar la información posible para estimar las habilidades. Se construyeron considerando dos grupos: el primero, integrado por las variables que están directamente en el modelo de regresión:

Variables que definen subpoblaciones de interés:

- Dominios primarios de PLANEA SEN: las entidades federativas
- Dominios secundarios de PLANEA SEN: tipos de escuelas que consideran sostenimiento y subsistema
- Sexo
- Edad

Variables que influyen sobre la habilidad de los estudiantes:

- La forma del examen que se le asignó.
- Estimación preliminar del promedio del rendimiento de cada estudiante en ambas asignaturas dentro de cada escuela. Es el promedio de los estimadores de máxima verosimilitud de ambas asignaturas para todos los demás estudiantes. Incluir las medias de rendimiento por escuela sirve para tomar en cuenta gran parte de la variación existente entre escuelas.

El segundo grupo, variables de condicionamiento, lo constituyen las variables que influyen indirectamente sobre la habilidad de los estudiantes. Dichas variables se construyen a partir de los cuestionarios de contexto de la siguiente forma:

1. Cada una de las variables de los cuestionarios de contexto A y B fueron recodificadas mediante el método *Deviance coding*. Es decir, se construyen variables indicadoras para cada una de las categorías de respuesta de cada una de las variables de los cuestionarios de contexto, a continuación se elige la categoría de referencia eligiendo la de mayor frecuencia estimada en la población; los valores de esa categoría son recodificados a -1.

2. Después se hizo un análisis de componentes principales para reducir la dimensionalidad. Se retuvieron las cargas factoriales que expliquen 95% de la varianza total de ambos cuestionarios de contexto. En general, la cantidad de factores que se construyen tiene que ser de al menos 200. Para bachillerato se consideraron 285 factores.

Una vez construidos los dos grupos de variables de condicionamiento, se estimaron los coeficientes de regresión de los que se obtuvieron dos grupos de coeficientes: uno para Lenguaje y Comunicación y otro para Matemáticas. Los coeficientes de regresión sirven para definir las medias de cada una de las subpoblaciones de interés.

Para terminar la estimación de los parámetros poblacionales, a la par de la estimación de los coeficientes de regresión, se estimó la matriz de varianza y covarianza fijando los parámetros de dificultad de los reactivos.

2.5 Extracción de valores plausibles

La distribución predictora necesaria para calcular (1), se puede aproximar numéricamente, una vez que fueron estimados los parámetros de dificultad de los reactivos, los coeficientes de regresión y la matriz de varianza y covarianza, aplicando el teorema de *Bayes* a $\boldsymbol{\theta}$,$\boldsymbol{X}$, $\boldsymbol{Y}$ y utilizando (4), (5) y (6).

$$p(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\delta}) = \boldsymbol{K_{\alpha\delta}}\, p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\delta}) p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\delta}) = \prod_{i=1}^{N} K_{i\alpha\delta}\, p(\boldsymbol{x}_i|\theta_i,\boldsymbol{\delta}) p(\theta_i|\boldsymbol{y}_i,\boldsymbol{x}_i,\boldsymbol{\alpha})$$

Donde,

$K_{i\alpha\delta} = \frac{1}{p(\boldsymbol{x}_i|\boldsymbol{y}_i,\boldsymbol{\alpha},\boldsymbol{\delta})}$ es la constante de normalización que depende de $\boldsymbol{\alpha}$ y $\boldsymbol{\delta}$ pero no de θ_i.

De esta distribución se extrajeron cinco vectores de valores plausibles por asignatura.

La confiabilidad final de PLANEA considera el modelo bidimensional, la inclusión de las variables de condicionamiento y los valores plausibles.

Asignatura	Confiabilidad bidimensional condicionada
Lenguaje y Comunicación	0.876
Matemáticas	0.884

2.5.1 Cálculo y cancelación del efecto del cuadernillo

El diseño matricial de las pruebas PLANEA SEN 2017 fue constituido por bloques incompletos balanceados, de modo que las estimaciones de los parámetros de los reactivos y de las habilidades de los estudiantes, que se obtienen en principio, no están influenciados por el efecto del cuadernillo. Sin embargo, en la práctica es común encontrar que aún hay influencia del cuadernillo debido a que los reactivos aparecen en diferentes posiciones dentro los cuadernillos.

Modelar el efecto de orden en términos de la posición de los reactivos dentro de un cuadernillo resultaría en un modelo muy complicado, por lo que se optó por usar el modelo implementado en EXCALE, el cual consistía en estimar el efecto de cada cuadernillo en la calibración, de la misma manera

que se estima la dificultad del cuadernillo; es decir, en la estimación de los parámetros de los reactivos, los efectos de los cuadernillos fueron incluidos en el modelo de medición para prevenir que se confundieran las dificultades de los reactivos y los efectos de los cuadernillos. El modelo empleado es de los llamados de facetas:

$$logit(p) = \theta - \delta_j - \tau_k$$

El parámetro del cuadernillo, τ_k, se definió en el mismo sentido de los parámetros de dificultad los reactivos, reflejando la dificultad de los cuadernillos.

Posteriormente a la estimación de las habilidades de los estudiantes con valores plausibles, los parámetros estimados de los cuadernillos fueron sumados a las estimaciones de habilidad de los estudiantes, con la finalidad de cancelar el efecto de cada uno de los cuadernillos.

2.6 Análisis de Datos con valores plausibles

Los valores plausibles pueden ser empleados para evaluar la ecuación (1), para cualquier estadístico Q, de la siguiente forma:

Usando cada uno de los vectores de valores plausibles se evalúa Q, como si los valores plausibles fueran los verdaderos valores de θ, cada una de estas evaluaciones se denotan con Q_m. Para $m = 1, \dots, M$ con $M = 10$.

Se estima la varianza de cada una de las evaluaciones de Q_m, cada una de estas estimaciones se denotan con U_m para $m = 1, \dots, M$.

La mejor estimación que se puede obtener de Q, a partir de los valores plausibles, es el promedio simple de los 10 valores obtenidos de los diferentes conjuntos de valores plausibles.

$$\hat{Q}_{pv} = \frac{1}{M}\sum_{m=1}^{M} Q_m$$

La estimación de la varianza total de $\hat{Q}_{pv}$ se hace sumando dos componentes.
Considere:

$$\overline{U} = \frac{1}{M}\sum_{m=1}^{M} U_m$$

$$B_M = \frac{1}{M-1}\sum_{m=1}^{M}\left(Q_m - \hat{Q}_{pv}\right)^2$$

La estimación de la varianza total de $\hat{Q}_{pv}$ es:

$$\hat{V}\left(\hat{Q}_{pv}\right) = \overline{U} + (1 + M^{-1})B_M$$

El primer término contiene la varianza debida al muestreo y el segundo, la varianza debida a que las estimaciones de las habilidades no se estiman con precisión.

Un intervalo de confianza a $(1-\alpha)\%$ para $\hat{Q}_{pv}$ está dado por:

$$\hat{Q}_{pv} \pm t_{\upsilon}\left(\frac{1-\alpha}{2}\right)\sqrt{\hat{V}(\hat{Q}_{pv})}$$

Donde t_{υ} es el cuantil de una distribución t con υ grados de libertad:

$$\upsilon = \left[\frac{f_M^2}{M-1} + \frac{(1-f_M)^2}{d}\right]^{-1}$$

Donde *d* son los grados de libertad para el estadístico calculado con los datos completos:

$$f_M = \frac{(1+M^{-1})B_M}{\hat{V}(\hat{Q}_{pv})}$$

2.6.1 Advertencia de posible sesgo en algunos análisis secundarios

Es común que se pretenda investigar asociaciones entre la variable latente de logro y alguna otra variable que no fue considerada en el condicionamiento. Sin embargo, esto no se debe de hacer, puesto que, si alguna variable no fue incluida como variable de condicionamiento, resulta imposible recobrar completamente el posible efecto o relación que tenga con la variable latente y, en consecuencia, las estimaciones que involucren a la variable no incluida en el condicionamiento, pueden resultar sesgadas en un grado desconocido. Este problema se encuentra ampliamente documentado en Mislevy (1991).

2.7 Definición de la escala basal

Una vez cancelado el efecto de los cuadernillos en los valores plausibles, es necesario aplicarles una transformación lineal con la finalidad de facilitar la comunicación de los resultados. Los valores plausibles se originan sobre la escala *logit*, así que la transformación lleva la media y la varianza de los valores plausibles a la nueva escala de PLANEA SEN, en la que se definió que la media fuera de 500 unidades y una desviación estándar de 100, para cada una de las asignaturas.
Se utilizó la siguiente transformación:

$$T = \frac{\sigma_1}{\sigma_0}\theta + \left[\mu_1 - \frac{\sigma_1}{\sigma_0}\mu_0\right]$$

En donde:
θ : Valor en la escala *logit*
σ_1 Es la desviación estándar de la escala de PLANEA SEN
μ_1 Es la media de la escala de PLANEA SEN

σ_0 Es la desviación estándar en la escala *logit* estimada con valores plausibles. Es decir, $\sigma_0 = \frac{1}{M}\sum_{m=1}^{M}\sigma_m$ es el promedio de las desviaciones estimadas para cada uno de los vectores de valores plausibles considerando el diseño de la muestra y los factores de expansión.
μ_0 Es la media en la escala *logit* estimada con valores plausibles. Es decir, $\mu_0 = \frac{1}{M}\sum_{m=1}^{M}\mu_m$ es el promedio de las medias estimadas para cada uno de los vectores de valores plausibles considerando el diseño de la muestra y los factores de expansión.

Con la definición de la escala base termina el proceso de escalamiento de resultados.

Apéndice B1. Tratamiento de los datos faltantes en los reactivos de logro

Los reactivos que se utilizaron en PLANEA para medir el logro de los estudiantes en las asignaturas de Lenguaje y Comunicación y Matemáticas fueron de opción múltiple. Para cada una de estas opciones de respuesta se asociaron los números del a 1 al 4.

En las evaluaciones de gran escala pueden ocurrir dos situaciones:

1. Los sustentantes respondieron una y sólo una de las posibles opciones de respuesta y se les asigno un puntaje tratándolos como reactivos dicotómicos; es decir, 1 si respondió correctamente y 0 cuando respondió incorrectamente.
2. Los sustentantes dejan de responder por alguna razón o responden más de una de las opciones de respuesta.

La segunda situación nunca es deseada, pues genera respuestas faltantes, sin embargo, es inevitable tener que lidiar con ella. Para PLANEA SEN, éstos y otros tipos de respuestas faltantes se consideraron como valores perdidos que recibieron un tratamiento distinto en los diferentes momentos del proceso de construcción de las escalas.

Como primer paso, se presentan las distintas categorías de respuestas faltantes consideradas en PLANEA SEN para los reactivos correspondientes al logro:

Respuesta omitida: ocurrió cuando un examinado tuvo el tiempo y la oportunidad de responder a un reactivo, sin embargo, no lo hizo. Se le asignó el código 9.
Respuesta múltiple: se consideró cuando un examinado respondió más de una de las opciones de respuesta. Se le asignó el código 8.
No administrada: las pruebas de logro se aplicaron mediante un esquema de rotación de bloques de reactivos siguiendo un esquema matricial, de tal forma que a cada examinado se le asignó un subconjunto a responder de toda la prueba. A las respuestas de los reactivos que no se les asignó para responder, se les consideró de esta clase y se les asignó el código 7.
No alcanzada: bajo las mismas condiciones de aplicación y tiempo, algunos examinados no alcanzaron a responder los reactivos del final de la sesión. A estas respuestas faltantes se les asigno el código 5.

Una vez identificados los distintos tipos de datos faltantes, se les asignó un puntaje diferente en cada una de las etapas de la construcción de las escalas. A continuación, se describe la asignación.

Puntajes para las respuestas faltantes

Para asignar los puntajes de las respuestas faltantes se siguió la estrategia sugerida por Adams, Wilson, y Wu (1997), en la que el tratamiento de las respuestas faltantes se hace diferenciado en un proceso de dos etapas.

Esta estrategia es la que usan en la actualidad las pruebas *TIMSS* en su proceso de escalamiento (Adams, Wilson y Wang, 1997).

En la primera etapa, en la que se hizo la estimación de los parámetros de dificultad de los reactivos bajo el modelo de Rasch (calibración), las respuestas no alcanzadas se trataron como no administradas, por lo que no se les asignó un puntaje.

La razón que describen Ludlow y O'Leary (1999) es la siguiente: si en este punto del proceso de estimación suponemos que las respuestas no alcanzadas fueran tratadas como incorrectas, ocurriría una sobreestimación de la dificultad de los reactivos ya que los estudiantes podrían contestar a algunas de estas preguntas correctamente, teniendo el tiempo y la oportunidad de responder a ellas.

En la segunda etapa, se consideraron fijos y conocidos los parámetros de dificultad de los reactivos y se hizo la estimación de las habilidades de los examinados. Las respuestas no alcanzadas se consideraron como respuestas incorrectas, por lo que se les asignó un puntaje 0.

Ludlow y O'Leary (1999) dicen que si suponemos que las respuestas no alcanzadas son tratadas como no administradas, en la fase de estimación de las habilidades los estudiantes podrían adoptar la estrategia de ir respondiendo únicamente los reactivos que está seguro de conocer la respuesta correcta, detenerse en el momento en el que ya no se sienta seguro y dejar el resto de la prueba en blanco. Como las respuestas no alcanzadas no serían tomadas en cuenta para la estimación de la habilidad, una gran parte de los alumnos podría obtener muy buenos resultados en la prueba. Por esta razón, en ambas etapas, las respuestas omitidas y múltiples se consideraron como respuestas incorrectas, y se les asignó un puntaje 0.

El mecanismo de administración de los reactivos mediante la rotación de los bloques bajo un diseño matricial, hace posible ignorar las respuestas no administradas, por lo que no se les asignó un puntaje, se les consideró *missing by design* (Boomsma, van Duijn, y Snijders, 2001).

Anexo 3. Reporte técnico de la construcción de índices y escalas de los datos de los cuestionarios de contexto

La Evaluación de Logro referida al Sistema Educativo Nacional (ELSEN) se realiza acompañada de un cuestionario de contexto diseñado específicamente para los alumnos. El Cuestionario para el alumno está integrado por una serie de reactivos orientados a conocer las características de orden personal, familiar, escolar y social de los alumnos, con la finalidad de obtener escalas que permitan analizar características no observables directamente.

El cuestionario de contexto para los alumnos consta de 120 reactivos. Un constructo que se midió con respuestas de este cuestionario es "Capital económico", para el cual se construye la escala. El procedimiento se encuentra en el punto 0 de este documento.

En los apartados 1 y 2 se describe a detalle cómo fueron construidos los siguientes índices:

> **Índices simples:** construidos sólo con transformaciones aritméticas o recodificaciones de uno o más reactivos. En este caso son siete: *Edad arriba de la típica, Edad en años cumplidos por categoría, Sexo, Condición indígena, Máximo nivel de estudios del padre, Máximo nivel de estudios de la madre y Máximo nivel de estudios alcanzado por los padres.*
> **Índices de escala:** construidos con el escalamiento de reactivos. Para estos índices, los puntajes de escala son estimaciones de rasgos latentes que se obtienen mediante el Modelo de la Teoría de respuesta al ítem (TRI) de crédito parcial (Masters, 1982) (MCP) para reactivos que tienen dos o más categorías de respuesta tipo ordinal.

En el apartado 0 se describe la metodología de escalamiento y ajuste del modelo. Esto incluye la descripción del Modelo de crédito parcial (MCP) y los criterios de ajuste: análisis factorial exploratorio y confirmatorio, revisión de las medidas de ajuste de los reactivos, análisis de consistencia interna de las escalas y análisis de la equidad de las escalas mediante el comportamiento diferencial de reactivos. En el apartado 4 se describen las recomendaciones de las características psicométricas de los reactivos dentro del modelo de medición de crédito parcial que permiten diagnosticar el comportamiento de reactivos dentro de una escala para garantizar su adecuado funcionamiento.

Para la construcción de índices simples de a) Edad y Sexo del alumno, y b) escala de Capital económico, revisar el *Apéndice* C1: Construcción de las variables de edad y sexo

en PLANEA EMS, 2017 y **¡Error! No se encuentra el origen de la referencia.**, respectivamente, los cuales hacen una descripción técnicamente amplia.

3.1 Índice simple

En esta sección se describe la construcción de los siete índices simples: *Edad arriba de la típica, Edad en años cumplidos por categoría, Sexo, Condición indígena, Máximo nivel de estudios del padre, Máximo nivel de estudios de la madre y Máximo nivel de estudios alcanzado por los padres.* Estos índices se

calculan mediante una transformación aritmética o recodificación a partir del registro de uno o más reactivos.

3.1.1 Edad y Sexo del alumno

El cálculo de la edad y sexo del alumno implicó un proceso minucioso derivado de considerar tres variables provenientes de dos fuentes de información: a) la CURP, tomada de la información oficial en el marco de muestreo y b) los años cumplidos al momento de la prueba y el mes de nacimiento provenientes de las respuestas del alumno. La descripción de este proceso se puede consultar en el Apéndice C1: Construcción de las variables de edad y sexo
en PLANEA EMS, 2017.

3.1.2 Condición indígena

Este índice simple reporta la respuesta del alumno a la pregunta del cuestionario de contexto "¿Alguno de tus padres habla una lengua indígena?". La estructura de los reactivos que la componen se muestra en la Tabla 1.

Tabla 1. Variables que componen el índice Condición indígena

Reactivo	Categorías de respuesta	Codificación
MA_2: ¿Alguno de tus padres habla una lengua indígena? Madre	D) Sí E) No F) No sé	D) 1 E) 2 F) 3
MA_3: ¿Alguno de tus padres habla una lengua indígena? Padre		

La construcción de este índice se conforma por las preguntas MA_2 y MA_3, quedando las categorías como se muestra en la Tabla 2.

Tabla 2. Construcción del índice Condición indígena

Condición indígena		MA_2 ¿Alguno de tus padres habla una lengua indígena? Madre		
		1 Sí	2 No	3 No sé
MA_3 ¿Alguno de tus padres habla una lengua indígena? Padre	1 Sí	1	1	1
	2 No	1	2	9
	3 No sé	1	9	9

"Condición indígena" es un índice simple compuesto por dos categorías válidas: 1: Sí y 2: No. Cuando el alumno reporta que al menos uno de sus padres habla una lengua indígena, entonces "Condición indígena" vale 1; si el alumno reporta que ni su madre ni su padre hablan una lengua indígena, entonces "Condición indígena" vale 2; si ninguna de estas condiciones se cumple, entonces "Condición indígena" es un valor perdido (9).

3.1.3 Nivel de estudios del padre y nivel de estudios de la madre

Estos dos índices simples reportan la respuesta del alumno a la pregunta: "¿Cuál es el nivel de estudios alcanzado por tu padre/madre (contesta aunque ya haya fallecido)?". El orden en el que se describen corresponde al mismo de las tablas de resultados públicos. La estructura de cada uno de los reactivos se muestra en la Tabla 3.

Tabla 3. Variables que componen los índices Nivel de estudios del padre y Nivel de estudios de la madre

Reactivo	Categorías de respuesta	Codificación
MA_111: ¿Cuál es el nivel de estudios alcanzado por tu padre (contesta aunque ya haya fallecido)?	G) No estudió H) Primaria I) Secundaria J) Nivel medio superior o carrera técnica K) Licenciatura L) Posgrado (especialidad, maestría, doctorado)	G) 1 H) 2 I) 3 J) 4 K) 5 L) 6
MA_110: ¿Cuál es el nivel de estudios alcanzado por tu madre (contesta aunque ya haya fallecido)?		

La construcción de cada uno de estos índices consistió en colapsar las dos últimas categorías de respuesta (5 y 6 donde 6 = 5) debido a la baja frecuencia reportada. De ese modo los índices se conforman como se muestra en la Tabla 4.

Tabla 4. Construcción de los índices Nivel de estudios del padre y Nivel de estudios de la madre

Índice simple	Categorías de respuesta	Recodificación
Máximo nivel de estudios del padre	F) No estudió G) Primaria H) Secundaria I) Nivel medio superior o carrera técnica J) Licenciatura o posgrado	F) 1 G) 2 H) 3 I) 4 J) 5
Máximo nivel de estudios de la madre		

3.1.4 Máximo nivel de estudios alcanzado por los padres

Éste es un índice simple que se conforma con los reactivos MA_110 y MA_111 descritos en la el punto 0. Para construir este índice se obtuvo el valor máximo reportado de ambos reactivos por alumno. Al final se colapsó la categoría 6 a la 5 para reportar la misma escala de categorías que los índices del punto 0. En la Tabla 5 se muestra la codificación del índice.

Tabla 5. Construcción del índice Máximo nivel de estudios alcanzado por los padres

Máximo nivel de estudios alcanzado por los padres		MA_111 ¿Cuál es el nivel de estudios alcanzado por tu padre (contesta aunque ya haya fallecido)?				
		1 No estudió	2 Primaria	3 Secundaria	4 Nivel medio superior o carrera técnica	5 Licenciatura o Posgrado
MA_110 ¿Cuál es el nivel de	1 No estudió	1	2	3	4	5
	2 Primaria	2	2	3	4	5

estudios alcanzado por tu madre (contesta aunque ya haya fallecido)?	3 Secundaria	3	3	3	4	5
	4 Nivel medio superior o carrera técnica	4	4	4	4	5
	5 Licenciatura o Posgrado	5	5	5	5	5

En cuanto a los valores perdidos, sólo en el caso en que ambos reactivos no reportaran un valor válido para la categoría se consideraba como perdido (9).

3.2 Índices de Escala

Los índices de escala se construyen a partir de un proceso de medición para variables latentes. En la medición intervienen tres elementos: un objeto de medición, un conjunto de números y un sistema de reglas. El objeto de medición puede ser una variable observable (como edad o sexo) o una variable latente, por ejemplo, habilidad para algo o actitud hacia algo. Cualquier variable latente se puede ver como un continuo de magnitudes que se incrementan en una dirección dada, digamos de izquierda a derecha si el continuo se representara con una línea recta. Una variable latente usualmente se define mediante indicadores observables como son las respuestas a los reactivos en un instrumento (Dimitrov, 2012). Las reglas sirven para asignar números o magnitudes al objeto de medición.

Los cuestionarios de contexto en PLANEA están conformados por reactivos con categorías de respuesta ordinal. Estas categorías de respuesta representan actitud, satisfacción o percepción hacia algo (por ejemplo, qué tan seguido ocurre algo). Para cada reactivo, los alumnos deben responder seleccionando una categoría que indique el nivel de su actitud, satisfacción o percepción en relación con la afirmación. La categoría seleccionada por el alumno se asocia con un valor numérico que representa el puntaje del alumno para ese reactivo.

Las reglas utilizadas para asignar magnitudes al objeto de medición conforman un proceso iterativo de ajuste del Modelo de crédito parcial (Masters, 1982) y los criterios de ajuste de la escala. Los criterios de ajuste de la escala incluyen el análisis de la dimensionalidad de las escalas con análisis factorial exploratorio y confirmatorio, el análisis de la consistencia interna de las escalas (Crocker, 2008) y la equidad de escalas mediante el análisis del comportamiento diferencial de reactivos (Wu, Adams, Wilson y Haldane, 2007).

A continuación se describen las escalas construidas para PLANEA 2015. Estas escalas son los objetos de medición que, en este caso, son rasgos o constructos latentes en los alumnos. Cada escala lleva el nombre del rasgo latente que se pretende medir. Se describen los reactivos que sirvieron como indicadores observables en cada caso y los puntajes que se asignaron a las categorías de cada reactivo.

3.2.1 Escala de Capital económico (CE)

Los siete reactivos que conforman esta escala proceden del cuestionario de contexto del alumno, los cuales se muestran en la Tabla 6. Para construir la escala, todos los reactivos han sido recodificados en sentido de ésta iniciando por el valor 0 hasta el valor k-1 categorías (Tabla 7). En el cuestionario, los reactivos Ma_117, Ma_118 y Ma_120 son dicotómicos y en sentido inverso a la escala. Televisor es un índice formado a partir de los reactivos Ma_115 y Ma_119 (Tabla 8).

Tabla 6. Reactivos que componen la escala Capital económico

Reactivo	Reactivo en el cuestionario	Etiqueta
TV	Ma_115 y Ma_119	Índice de televisor: número de aparatos, considerando servicio de cable.
MA_114	Ma_114	Del siguiente listado de bienes, señala cuántos hay en tu casa. Computadora
MA_116	Ma_116	Del siguiente listado de bienes, señala cuántos hay en tu casa. Automóvil
MA_117	Ma_117	En tu casa, ¿cuentas con los siguientes bienes y servicios? Teléfono convencional (fijo)
MA_118	Ma_118	En tu casa, ¿cuentas con los siguientes bienes y servicios? Horno de microondas
MA_120	Ma_120	En tu casa, ¿cuentas con los siguientes bienes y servicios? Internet

Tabla 7. Categorías de las variables que componen la escala Capital económico

Variables	Categorías
TV construida a partir de: Ma_115 y Ma_119	0: No tiene televisor, 1: Tiene al menos un televisor (sin servicio de cable o vía satélite), 2: Tiene 1 o 2 televisores y servicio de cable o vía satélite, 3: Tiene 3 o más televisores y servicio de cable o vía satélite
MA_114, MA_116	0: Ninguno, 1: 1, 2: 2 o 3, 3: 4 o más
MA _117, MA _118, MA_120	0: No, 1: Sí

Tabla 8. Construcción del índice Televisor

<table>
<tr><th colspan="2" rowspan="2">(TV) Índice Televisor (número de aparatos, considerando servicio de cable)</th><th colspan="4">MA_119 En tu casa, ¿cuentas con los siguientes bienes y servicios? Televisión por cable o vía satélite</th></tr>
<tr><th>(1) Sí</th><th>(2) No</th><th>(8) Respuesta múltiple</th><th>(9) Respuesta omitida</th></tr>
<tr><td rowspan="6">MA_115 Del siguiente listado de bienes, señala cuántos hay en tu casa. Televisor</td><td>(1) Ninguno</td><td colspan="4">(0) No tiene televisión</td></tr>
<tr><td>(2) 1</td><td rowspan="2">(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3" rowspan="2">(1) Tiene al menos un televisor</td></tr>
<tr><td>(3) 2</td></tr>
<tr><td>(4) 3</td><td rowspan="2">(3) Tiene 3 o más televisores y servicio de cable o vía satélite</td></tr>
<tr><td>(5) 4 o más</td></tr>
<tr><td>(8) Respuesta múltiple</td><td>(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3">( ) Respuesta omitida</td></tr>
</table>

El procedimiento de construcción y validación de la escala se puede consultar en el **¡Error! No se encuentra el origen de la referencia.**.

3.3 Metodología de escalamiento y ajuste del modelo

El escalamiento está basado en los procedimientos de PISA (OECD, 2014; OECD, PISA, 2006; Technical Report, 2009).

3.3.1 Procedimiento de escalamiento

El escalamiento o construcción de escalas de medición del constructo se realizó bajo el Modelo de crédito parcial (MCP) de Masters (1982) perteneciente a la familia de modelos Rasch utilizados en la Teoría de respuesta al ítem. El MCP se utiliza para modelar probabilidades entre categorías adyacentes de respuestas al reactivo. Tradicionalmente, la escala de respuestas al reactivo es de cuatro categorías a las cuales se asignan puntuaciones 0, 1, 2 y 3, donde valores mayores indican mayor cantidad del constructo, atributo o rasgo latente que se va a medir (Wilson, 2005).

Modelo de crédito parcial

A continuación se presenta el planteamiento del Modelo de crédito parcial basado en Van der Linden y Hambleton (1997). El Modelo de crédito parcial permite analizar respuestas que se registran en dos o más categorías ordenadas; tiene la característica de contar con parámetros separados de personas y reactivos que permiten hacer comparaciones. Todos los parámetros del modelo son *localizaciones* sobre una variable o rasgo latente.

Los parámetros del reactivo i $\delta_{i1}, \delta_{i2}, \dots, \delta_{im_i-1}$ para cada categoría representan la posición sobre la escala del rasgo latente medido en la que un alumno tiene la misma probabilidad de responder en la categoría x y en la categoría $x-1$ (ie, $P(x_{ni}=x|\theta_n,\delta_{ix}) = P(x_{ni}=x-1|\theta_n,\delta_{ix})$). Los valores de los parámetros de reactivo no necesariamente se encuentran ordenados con respecto al puntaje. En evaluaciones de logro δ_{ik} se denomina *dificultad* del reactivo.

El parámetro para alumnos (o cantidad de rasgo sobre la escala *logit*) θ_n es la localización modelada del alumno n sobre la escala del rasgo latente medido. En evaluaciones de logro θ_n se denomina *habilidad* del alumno.

Todos los parámetros son ubicaciones sobre la escala del rasgo latente medido y pueden utilizarse para mapear el significado cualitativo de la variable y para interpretar los parámetros de los alumnos.

Otra manera de definir los parámetros de los reactivos del modelo es $\delta_{ik} = \delta_{i.} + \tau_{ik}$, donde $\delta_{i.}$ es el valor esperado de las localizaciones $\delta_{i1}, \delta_{i2}, \dots, \delta_{im_i-1}$ para el reactivo i y $\delta_{i.} + \tau_{ik}$ es el umbral de localización o umbral de Rasch-Andrich, donde τ_{ik} es un umbral que compensa la ubicación de las categorías respecto a $\delta_{i.}$. Cuando el modelo restringe a las localizaciones como combinación linealmente independiente, esto es, $\sum \delta_{ik} = 0$, entonces los umbrales de localización también son linealmente independientes y entonces $\sum \tau_{ik} = 0$.

Sea el i-ésimo reactivo con puntajes de categoría $0,1, \dots, m_i$, la probabilidad de que el n-ésimo alumno obtenga el puntaje x, con localización θ_n, está dado por:

$$P(x_{ni} = x|\theta_n, \delta_{ik}) = \frac{exp \sum_{k=0}^{x}(\theta_n - \delta_{ik})}{\sum_{h=0}^{m_i} exp \sum_{k=0}^{h}(\theta_n - \delta_{ik})}$$

En este modelo:

Las categorías de cada reactivo están ordenadas de menor a mayor puntaje y se descomponen en pares de categorías adyacentes. Se aplica un modelo dicotómico para cada par, de manera que el alumno tiene mayor cantidad del rasgo latente cuando obtiene mayor puntaje.

La probabilidad asignada a cada categoría es positiva, esto es, la probabilidad de que un alumno alcance la puntuación correspondiente a esa categoría es mayor a cero y la probabilidad total de cada reactivo es la suma de todas las probabilidades de categoría. Esto indica que para que un alumno alcance la probabilidad de obtener k puntos más que $k-1$ puntos es porque ha superado la etapa de obtener $k-1$ puntos. Lo que se dice es que este alumno "ha pasado a través de" la etapa de $k-1$ puntos. Pero además también ha pasado a través de la etapa de los puntos $k-2, k-3, \ldots, 0$.

Cuando se estiman los parámetros para los reactivos y cumplen con los criterios de validación que se describen en el apartado 0, se utiliza estimación de máxima verosimilitud ponderada para obtener puntajes individuales para los alumnos. Estos puntajes se conocen como wle (por sus siglas en inglés) y se obtienen mediante el software ACER ConQuest®.

En la base de datos final, los parámetros estimados wle y $\hat{\delta}_{ik}$ se presentan trasladados. Este traslado se realiza con el procedimiento que se describe en el apartado 0.

Estadísticas de ajuste al Modelo de crédito parcial

Correlación punto biserial

La correlación punto biserial para la categoría k en el reactivo i se calcula con una variable indicadora y_{ikn} que vale 1 para los alumnos que anotan el puntaje de la categoría k en el reactivo i. La correlación punto biserial se calcula entre el conjunto de valores y_{ikn} y el correspondiente s_n. Donde s_n se define en ConQuest® como el puntaje crudo para un alumno, el cual consiste de la suma del puntaje alcanzado por éste dividido por el puntaje máximo posible que el alumno podría haber alcanzado. Sólo aquellos casos que respondieron al reactivo son incluidos en el cálculo (Wu, Adams, Wilson y Haldane, 2007, p. 149).

Estadísticos de ajuste basados en residuales

Los estadísticos de ajuste basados en residuales son una forma de índices de ajuste para reactivos y para alumnos, derivados de las diferencias entre el puntaje observado y el puntaje esperado del reactivo (residuales) por cada reactivo y por cada alumno. Wright (1977) propuso varios de estos estadísticos basados en los residuales estandarizados.

El residual estandarizado es la diferencia del puntaje obtenido por el alumno a un determinado reactivo (puntaje observado) menos el puntaje esperado dividido por la varianza del puntaje (Wu y Adams, 2007).

Outfit (cuadrado medio del error). Es la suma de los cuadrados de los residuales estandarizados de todos los alumnos dividida por el número total de alumnos, se puede demostrar que el valor esperado de este estadístico es uno.

Infit (cuadrado medio del error ponderado). Es la suma del cuadrado del residual ponderado por la varianza de la respuesta al reactivo, se puede demostrar que el valor esperado de este estadístico es uno.

De acuerdo con Boone, Staver y Yale (2014) cuando no hay valores atípicos es indistinto valorar el outfit y el infit. En nuestro caso se valoró el infit.

Rango de valores aceptables para el infit

Existen diversos autores que han establecido rangos de aceptabilidad para el ajuste basado en los residuales estandarizados, para fines de interpretación en este documento el tipo de ajuste y rango a considerar es el reportado por Wilson (2005), quien establece que si bien no existe un límite absoluto de lo que es un buen valor cuadrado medio ponderado (infit), 0.75 es un valor inferior razonable, así como 1.33 lo es como valor superior.

Es importante notar que estos estadísticos de ajuste no indican qué tan cercana está la curva observada de la esperada, sino más bien que la pendiente de la curva observada es la misma que la de la esperada o teórica, es decir, el cuadrado medio de ajuste no mostrará la falta de ajuste si la curva característica observada del reactivo está cerca o lejos de la curva característica esperada del reactivo (Wu y Adams, 2007).

Traslado de la escala

Las mediciones de las escalas se reportan con media 50 y desviación estándar 10 para evitar el uso de números negativos y disminuir la cantidad de decimales. Esta transformación de la escala es posible ya que cualquier transformación lineal preserva la interpretación de la probabilidad de los valores *logits* (Wilson, 2005). A continuación se describe el procedimiento de traslado de escala basado en la descripción de Wright y Stone (1979).

Los valores originales de la escala se trasladan mediante una trasformación lineal de tipo:

$$y = \alpha + \gamma x$$

En la cual x está en la escala *logit* original mientras que y está en la escala que se denomina trasladada donde α representa el origen de la escala trasladada que es el valor de y, donde la escala *logit* original x vale 0, γ denota una constante de espaciamiento que determina las unidades de la escala trasladada. Si denotamos a y como B para los valores de los alumnos en la escala trasladada y como D para los valores de los reactivos en la escala trasladada, θ y δ son los valores originales para alumnos y reactivos respectivamente. Por lo que la ecuación de la recta para cada valor en la escala trasladada sería:

$B = \alpha + \gamma\theta$ para alumnos
$D = \alpha + \gamma\delta$ para reactivos

Las desviaciones estándar (De) de las mediciones para los alumnos y los reactivos cumplen que:
$De(B) = \gamma De(\theta)$

$$De(D) = \gamma De(\delta)$$

Se realiza una transformación preliminar utilizando los valores de la media m y desviación estándar s de la escala *logit* original para obtener valores θ' y δ' con media 0 y desviación estándar 1 de la forma:

$$\theta' = (\theta - m)/s$$
$$\delta' = (\delta - m)/s$$

Se toma el valor de origen $\alpha = 50$ y de espaciamiento $\gamma = 10$ para, finalmente, obtener los valores correspondientes a la escala trasladada con las siguientes ecuaciones:

$$B = \alpha + \gamma\theta' = \alpha + \gamma(\theta - m)/s$$
$$D = \alpha + \gamma\delta' = \alpha + \gamma(\delta - m)/s$$

Con esto se tiene que la escala trasladada tiene media 50 y desviación estándar 10.

3.3.2 Análisis de la dimensionalidad de las escalas con análisis factorial exploratorio y confirmatorio

El análisis factorial es una técnica estadística que determina la estructura subyacente entre las variables en el análisis. Es decir, es factible identificar grupos de variables dependiendo de sus correlaciones, cada conjunto de variables identificadas se denomina factor.

Un factor es una combinación lineal de las variables observadas. Los factores representan las dimensiones (constructos) subyacentes que resumen o representan a las variables observadas (Hair, Black, Babin y Anderson, 2009).

El análisis factorial se puede llevar a cabo desde dos perspectivas: análisis factorial exploratorio (AFE) y análisis factorial confirmatorio (AFC). El AFE permite identificar la existencia de factores en el conjunto total de reactivos de interés, suele ser utilizado cuando los constructos y los reactivos asociados con ellos no cuentan con un vasto historial de investigación; y el AFC se usa cuando se han identificado factores que se pueden asociar a un constructo de interés y se desea probar la unidimensionalidad de éstos.

Las escalas de la prueba PLANEA 2015 (Habilidades para la convivencia escolar, Recursos asociados al bienestar familiar y Oportunidades de aprendizaje fuera de la escuela) fueron sometidas a estos análisis con el fin de corroborar que los reactivos candidatos a formar las escalas fueran unidimensionales.

3.3.3 Análisis factorial exploratorio

El análisis factorial exploratorio es útil para encontrar una forma de concentrar la información obtenida de las respuestas de los reactivos (variables observadas) en un número pequeño de factores (variables latentes). En este proceso se suele perder información (variabilidad), es por ello que también se busca que haya una pérdida mínima de información, es decir, que los factores logren recuperar la mayor variabilidad de los reactivos.

El modelo empleado para medir observaciones en relación a k factores, es el Modelo lineal de análisis factorial (MLAF) que es:

$$x = \Lambda_x \xi + \epsilon$$

Donde x es el vector de variables observadas (reactivos) cuya dimensión es $q \times 1$, donde q es el número de reactivos/preguntas en el análisis para *N* estudiantes. La matriz Λ_x llamada matriz de cargas factoriales, cuya dimensión es $q \times k$, contiene las cargas/coeficientes de regresión para cada factor, las entradas de las matriz son λ_{ij} que representan la carga al factor i del reactivo j. El vector ξ de dimensión $k \times 1$, es el vector de factores. Finalmente ϵ es un vector $q \times 1$ de variables únicas que contienen tanto el error de medición como el error específico.

Los supuestos a cumplir en torno a este modelo son:

$$E(\xi) = 0$$
$$E(\epsilon) = 0$$
$$cov\,(\xi, \epsilon) = 0$$

Donde *cov*(·,·) es el operador covarianza. Bajo estos supuestos, la matriz de covarianzas de los datos observados puede ser escrita en la forma de la ecuación fundamental de análisis factorial de modo que:

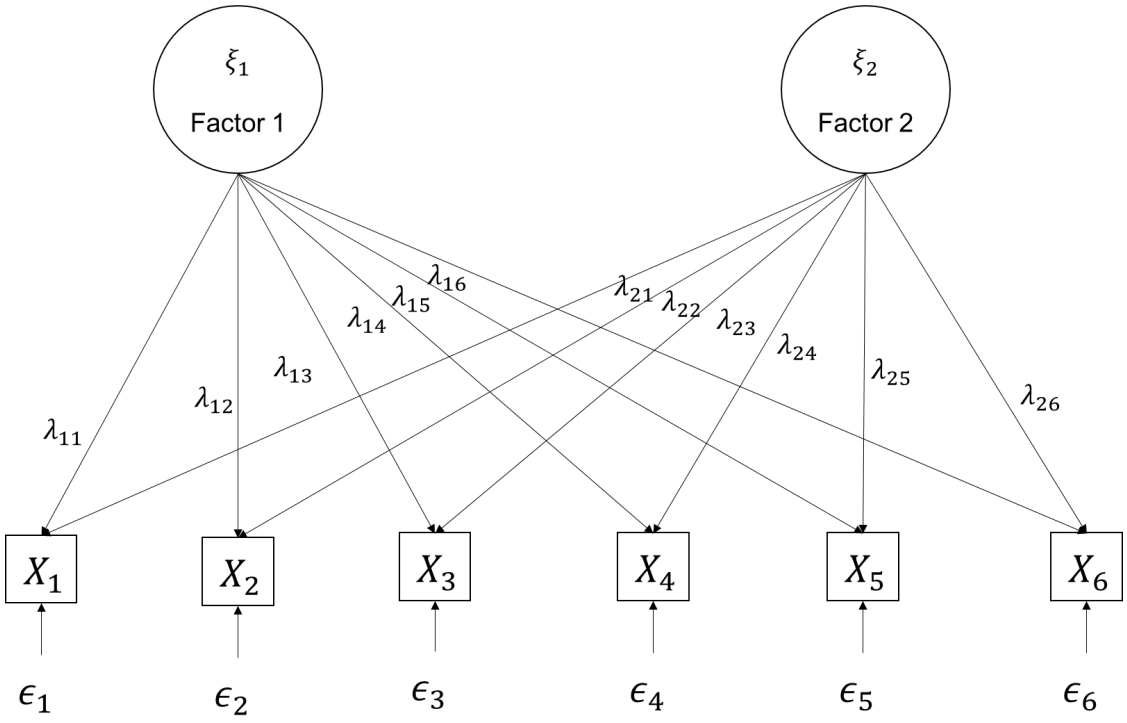
$$\sum = cov(\boldsymbol{x}\boldsymbol{x}')$$
$$= \Lambda_x E(\xi\xi')\Lambda'_x + E(\epsilon\epsilon')$$
$$= \Lambda_x \Phi \Lambda'_x + \Theta_\epsilon$$

Donde Σ es la matriz de covarianzas poblacional *q* x *q*, Φ es la matriz *k* x *k* de varianzas y covarianzas de los factores y Θ_ϵ es una matriz diagonal *q* x *q* de varianzas únicas.

El modelo supone un número k de factores, por lo tanto para identificar el valor apropiado en un análisis factorial exploratorio se suele utilizar un gráfico de sedimentación (Cattell, 1966), en éste se grafican ordenados los autovalores de la matriz de correlación. Se considera como número óptimo de factores el número de autovalores mayores a uno (Thompson, 2004). También se toma en cuenta el porcentaje de varianza explicada acumulada conforme se aumenta el número de factores, un 60% es satisfactorio (Hair, Black, Babin y Anderson, 2009). El porcentaje de varianza explicada en cada factor se calcula como el porcentaje que cada autovalor aporta a la suma de autovalores.

En la Figura 1 se muestra gráficamente el modelo empleado en el análisis factorial exploratorio para el caso de dos factores. Una vez que se ha identificado el número de factores se reconoce qué reactivos se encuentran asociados a cada factor. Para ello se analizan las cargas factoriales de la relación del reactivo con cada factor. Hair, Black, Babin y Anderson (2009) recomiendan conservar aquellas relaciones con carga factorial mayor o igual a 0.5 y cargas factoriales entre 0.3 y 0.4 son mínimamente aceptables.

Figura 1. Modelo factorial exploratorio con dos factores y seis reactivos.



Adaptado por Thompson (2004: 38).

Cuando no es sencilla la interpretación de la asociación de los reactivos con los factores, Hair, Black, Babin y Anderson (2009) sugieren emplear un método de rotación para lograr soluciones de factor simples y teóricamente con mayor interpretación; el efecto final de la rotación de la matriz de factores es redistribuir la varianza de los factores anteriores (no rotados) a los posteriores (rotados). En otras palabras, la rotación de factor implica mover los ejes de medición del factor de las ubicaciones de las variables medidas en el espacio factorial, de modo que la naturaleza de los constructos subyacentes se hace más obvia para el investigador; cabe mencionar que la rotación no es posible cuando un solo factor es extraído (Thompson, 2004). En el caso de las escalas de los cuestionarios A y B de PLANEA 2015 no se hizo rotación alguna porque el objetivo era probar la unidimensionalidad de las escalas, es decir, el comportamiento de un factor.

En la literatura existen diferentes métodos de rotación entre los que destacan rotaciones ortogonales[2] y oblicuas[3] (Hair, Black, Babin y Anderson, 2009; Thompson, 2004; Raykov y Marcoulides, 2011;

[2] Algunos ejemplos de rotaciones ortogonales son Varimax, Quartimax, Equimax.
[3] Algunos ejemplos de rotaciones oblicuas son Geomin, Oblimin, Promax.

Asparouhov y Muthén, 2009). En general, no hay una regla específica para utilizar algún tipo de rotación (Hair, Black, Babin y Anderson, 2009); los autores Asparouhov y Muthén (2009: 429) argumentan que no hay una razón estadística para preferir un criterio de rotación sobre otro, la decisión e interpretación de los resultados está enteramente en las manos del investigador. Sin embargo, las rotaciones ortogonales son las más utilizada cuando el propósito es una estructura simple[4]; si el propósito de la investigación es indagar sobre las correlaciones entre las constructos vinculados con los factores, se suele usar rotaciones oblicuas (Thompson, 2004). Es por esta razón que Raykov y Marcoulides (2011) indican que una rotación oblicua tiende a producir resultados más valiosos en investigaciones de comportamiento y sociales.

Harrington (2009: 10) recomienda que si una nueva medida de constructo está siendo desarrollada con un fuerte marco teórico, entonces se puede omitir el análisis factorial exploratorio y continuar con el análisis factorial confirmatorio.

En el caso de las escalas construidas con la información de los cuestionarios de PLANEA 2015, el análisis exploratorio se elaboró usando el software Stata (StataCorp, 2011) y Mplus (Muthén y Muthén, 2015) para calcular las estimaciones del modelo antes descrito y verificar que se cumplieran las condiciones antes mencionadas.

3.3.4 Análisis factorial confirmatorio

Después de verificar, en el análisis factorial exploratorio, que la estructura de los reactivos fuera acorde con las escalas, se elaboró un análisis factorial confirmatorio. Este análisis se realizó únicamente en las escalas para saber si se cumple el supuesto de unidimensionalidad. Previamente se prepararon los datos de manera que las categorías estén en sentido creciente hacia la escala, que no haya frecuencias menores a 5% en alguna de ellas y, si las hay, éstas categorías se colapsan con la categoría aledaña.

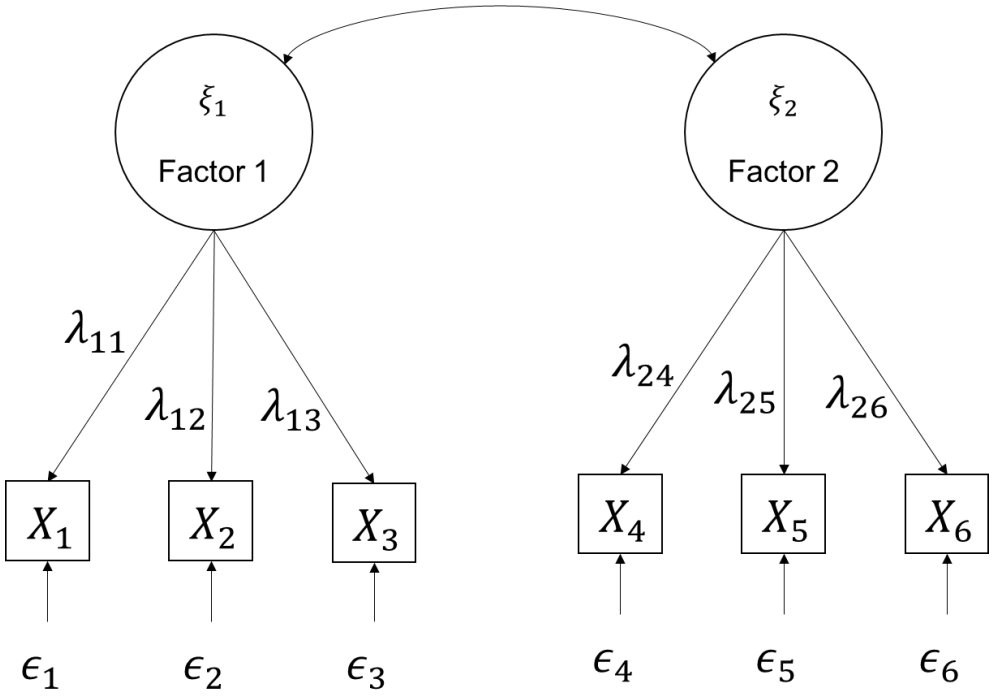
El modelo matemático empleado en este análisis es el mismo que el anterior (MLAF), salvo que existen restricciones en la matriz Λ_x (debido a que el número y las restricciones impuestas en la matriz asociada al modelo factorial son más de k^2). Las restricciones impuestas en los elementos de Λ_x reflejan una hipótesis *a priori* de estructura simple dada por el investigador Crocker (2008). En la Figura 2 se pueden apreciar de forma gráfica las restricciones, esto es, se restringe a que cada reactivo esté relacionado a un solo factor, teniendo carga factorial cero el resto de las relaciones. Lo anterior significa que el modelo para cada escala puede ser planteado como MLAF de un solo factor. Cabe señalar que, a diferencia del análisis exploratorio, no es posible rotar el modelo restringido porque destruiría el posicionamiento de las restricciones y, en consecuencia, la hipótesis bajo estudio (Kaplan, 2000). En este sentido, Brown (2015: 37) argumenta que en un AFC no es necesario rotar porque un reactivo está especificado para un solo factor.

Las escalas obtenidas de PLANEA 2015 fueron evaluadas calculando las estimaciones del modelo teórico y se verificó que cumplieran con las recomendaciones hechas por varios autores en la literatura. A continuación, se describen dichas recomendaciones e interpretaciones para las estimaciones del modelo.

[4] Una sola carga factorial alta para cada variable en sólo un factor (Hair, Black, Babin y Anderson, 2009).

El software empleado para obtener los resultados del modelo suelen mostrar las estimaciones de las cargas factoriales y correlaciones, así como las estimaciones estandarizadas de las mismas; estas últimas corresponden a las estimaciones obtenidas tomando en cuenta tanto los valores estandarizados de la escala (factor) como de los reactivos, es decir reescala los valores de la escala (factor) y del reactivo para que su varianza sea igual a uno (Wang y Wang, 2012: 39).

Figura 2. Modelo factorial confirmatorio con dos factores, cada uno de ellos con tres reactivos.



Adaptado por Thompson (2004: 38).

La estimación estandarizada de las cargas factoriales representa la relación del reactivo con la variable latente; por ejemplo, si la carga factorial de un reactivo tuviera valor de 0.8 para una escala, un incremento estandarizado en el valor de la escala está asociado a un incremento estandarizado de 0.8 en el reactivo. El cuadrado de la estimación estandarizada de la carga factorial es la proporción de varianza del reactivo, que es explicada por la escala, también suele llamarse a este valor la comunalidad del reactivo (Brown, 2015). Siguiendo con el ejemplo, una carga factorial del 0.8 representa 64% (0.8^2=0.64) de la variabilidad del reactivo está siendo explicada por la escala. Wang y Wang (2012) señalan que, convencionalmente, una carga factorial con valor de 0.3 es considerada como un valor de corte, aunque también suele usarse el valor de 0.4. Por otro lado, Hair, Black, Babin y Anderson (2009) sugieren como regla de dedo que la carga factorial estandarizada estimada debería ser mayor a 0.5 e idealmente mayor a 0.7.

Los residuos de correlación (o de covarianza) son la diferencia entre la matriz de correlación (o covarianza) implicada por el modelo y la matriz de correlación observada. Kline (2005) tiene como regla de dedo que un residuo de correlación mayor a 0.10 sugiere que el modelo estimado no explica adecuadamente la correlación entre los elementos del modelo; cuando el modelo es adecuado, los

residuos de correlación tendrán una distribución normal. Por su parte, Hoyle (1995) menciona que si el promedio de los valores absolutos de los residuos de correlación es de 0.02, el modelo es "bueno", esto se debe a que el promedio de error en las correlaciones es de 0.02 (aun existiendo correlaciones altas y bajas). Hoyle (1995) también argumenta que si el valor más grande en los residuales de correlación es pequeño, es decir 0.10, el modelo es sólo marginalmente inadecuado para algunas variables; si la mayor discrepancia entre la correlación implicada por el modelo y la matriz de correlación observada es grande, es decir 0.4, claramente el modelo no explica algunas de las correlaciones.

Una vez que se obtienen las estimaciones del modelo en un análisis factorial confirmatorio, se evalúa el modelo, dicho de otra manera: se verifica la bondad de ajuste del modelo. Al efecto se han desarrollado índices que reflejen la discrepancia entre la matriz de covarianzas muestral y la matriz de covarianzas implicada por el modelo con las cargas factoriales y correlaciones estimadas (Lei y Wu, 2007). El primer índice en ser desarrollado fue la estadística χ^2 (Brown, 2015), sin embargo este índice se usa poco debido a su sensibilidad respecto al tamaño de muestra (Lei y Wu, 2007; Wang y Wang, 2012; Brown, 2015). En seguida se enlistan los índices alternativos, usados para evaluar la bondad de ajuste del modelo.

Índices incrementales o comparativos. Índice Tucker-Lewis (TLI, por sus siglas en inglés) o índice de ajuste no normado (NNFI), índice de ajuste comparativo (CFI, por sus siglas en inglés: *Comparative Fit Index*) y el índice de ajuste normado (NFI, por sus siglas en inglés: *Normed Fit Index*) compara el modelo estimado con el modelo nulo[5]. El rango de los índices está entre cero y uno, valores grandes indican mejora del modelo estimado respecto al modelo nulo; valores mayores o igual a 0.9 son generalmente aceptados como indicadores de un buen ajuste. Se ha sugerido incrementar el punto de corte a 0.95 (Lei y Wu, 2007; Wang y Wang, 2012; Brown, 2015).

Índices de ajuste absoluto. Raíz cuadrada media residual estandarizado, (SRMR por sus siglas en inglés: *Standardized root mean square residual*), puede ser visto como el promedio entre la discrepancia de las correlaciones observadas y las correlaciones predichas por el modelo. Un criterio para la indicación de un buen ajuste del modelo es SRMR menor o igual que 0.08, valores entre 0.08 y 0.10 son aceptables (Lei y Wu, 2007; Wang y Wang, 2012; Brown, 2015). Media cuadrática del error de aproximación[6] (RMSEA, por sus siglas en inglés: *Root mean square error of approximation*), es un índice de "error de aproximación" porque evalúa el grado con el que el modelo se ajusta razonablemente bien a la población. Valores de RMSEA menores o iguales a 0.06 indican un buen ajuste del modelo (Brown, 2015); algunos autores citan otros rangos como Wang y Wang (2012) y Kaplan (2000): igual a cero indica ajuste perfecto, menor a 0.05 indica ajuste cercano, entre 0.05 y 0.08 indican ajuste razonable, de 0.08 a 0.10 ajuste mediocre y mayor a 0.10 mal ajuste.

Índices de criterio de información. Son criterios estadísticos empleados para la comparación de modelos, esto es porque en ocasiones se desea un modelo sobre un modelo alternativo que pudiera tener más o menos correlaciones o cargas factoriales a estimar. Los criterios están en función del modelo de máxima verosimilitud (valores de verosimilitud cercanos a uno implican valor del criterio bajo, valores de verosimilitud cercanos a cero implican valor del criterio alto) más una penalización en función del número de observaciones y del número de correlaciones y cargas factoriales a estimar (Wang y Wang,

[5] Modelo en el cual las covarianzas entre todos los indicadores se fijan como cero (Brown, 2015).

[6] Brown (2015: 71) lo clasifica como un índice de corrección de parsimonia, los cuales incorporan una función de penalización por modelo pobre en parsimonia.

2012). Criterio de información de Akaike (AIC, por sus siglas en inglés: *Akaike's information criterion*), criterio de información Bayesiana (BIC, por sus siglas en inglés: *Bayesian information criterion*) y criterio de información Bayesiana ajustado al tamaño de muestra (ABIC, por sus siglas en inglés: *Adjusted Bayesian information criterion*). El modelo con menor medida del criterio de información tiene un mejor ajuste a los datos en relación a los modelos alternativos (Wang y Wang, 2012; Brown, 2015).

Si un modelo no cumple adecuadamente con las recomendaciones de un buen ajuste, este puede modificarse añadiendo o removiendo correlaciones o cargas factoriales considerando el marco teórico y la interpretación de las modificaciones (Harrington, 2009; Wang y Wang, 2012; Brown, 2015).

Los índices de modificación (MI, por sus siglas en inglés: *Modification index*), también llamados multiplicadores de Lagrange, son cantidades calculadas para cada carga factorial y correlación fija del modelo, es decir, que no fueron estimados dentro del modelo debido a que se fija su valor a cero. Los MI reflejan una aproximación de que tanto la estadística χ^2 decrecerá si dicha carga factorial o correlación fuera estimada y no fija como en el modelo original (Hair, Black, Babin y Anderson, 2009; Brown, 2015). Brown (2015) indica que el MI puede conceptualizarse con una estadística χ^2 con 1 grado de libertad, por lo tanto valores mayores o iguales a 3.84[7] sugieren que el estimar libremente la carga factorial o correlación mejorará significativamente el ajuste del modelo. Sin embargo, el MI es sensible al tamaño de muestra (Hair, Black, Babin y Anderson, 2009; Brown, 2015), por lo cual se sugiere usar estos índices en conjunción con los *parámetros esperados de cambio* (EPC, por sus siglas en inglés: *expected parameter change*) debido a que estos últimos son útiles para muestras grandes (Kaplan, 1989; Wang y Wang, 2012; Millsap y Olivera-Aguilar, 2012; Whittaker, 2012; Brown, 2015).

Los EPC son valores que indican el valor esperado de la estimación de la carga factorial o correlación, si ésta hubiera sido estimada, análogo a estos valores existen los *parámetros esperados de cambio estandarizados* (SEPC, por sus siglas en inglés). En Brown (2015) se cita el trabajo de Jöreskog (1993), quien sugiere que se estime libremente la carga factorial o la correlación, es decir, que se haga la modificación sobre aquella cuyo MI es el más grande (y el EPC) si la estimación puede ser interpretada sustancialmente; en caso de no existir base sustancial se considera el siguiente más grande en MI y EPC para darle el mismo tratamiento.

En el caso de las escalas de PLANEA 2015, el análisis fue realizado con el programa EQS (Bentler y Wu, 2014) tomando en consideración los criterios antes mencionados.

3.3.5 Análisis de la consistencia interna de las escalas mediante el coeficiente de confiabilidad Alpha de Chronbach

La confiabilidad de las escalas se midió con el coeficiente Alpha de Chronbach. Este coeficiente se estima mediante la expresión:

$$\widehat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\sigma_x^2}\right)$$

[7] Una distribución χ^2 con un grado de libertad es igual a la distribución normal estandarizada al cuadrado, por lo tanto el punto 1.96^2=3.84 es el punto crítico con un $\alpha = 0.05$.

Donde k es el número de reactivos en la prueba, σ_i^2 es la varianza del reactivo i y σ_x^2 es la varianza total de la prueba (Crocker, 2008).

Este coeficiente es cercano a uno cuando la suma de las varianzas de los reactivos es más pequeña respecto a la varianza total de la prueba. En el Cambridge Dictionary of Statistics (Everitt y Skrondal, 2010) la interpretación sugerida para este coeficiente es: menor a 0.60 inaceptable, (0.60, 0.65) indeseable, (0.65, 0.70) mínimamente aceptable, (0.70, 0.80) bueno, (0.80, 0.90) muy bueno y mayor a 0.90 considere acortar la escala reduciendo el número de reactivos (*Statistical Evaluation of Measurement Errors: Desgin and Analysis of Reliability Studies*, 2004). En general, el uso que se le da al instrumento determina los valores aceptables para este coeficiente.

La confiabilidad también se calcula eliminando un reactivo a la vez. Si el valor del alfa aumenta considerablemente al eliminar un reactivo, se considera no incluir tal reactivo en la escala final.

Además, se calcula la correlación reactivo-total, que es el coeficiente de correlación de Pearson de un reactivo individual con el total de la escala calculada sumando los reactivos restantes. Una regla de pulgar indica que los reactivos que correlacionan abajo de 0.20 deben ser descartados (Streiner y Norman, 2010; Everitt y Skrondal, 2010). Los reactivos con correlaciones negativas podrían estar en sentido inverso a la escala. Se deben reacomodar sus categorías de manera que a mayor valor de la categoría de respuesta, indique mayor cantidad del rasgo latente en el alumno. Es decir que los puntajes de las categorías de respuesta aumenten en el sentido de la escala. Todos los cálculos anteriores se realizaron con el programa SPSS Statistics® 23.

3.3.6 Análisis de la equidad de las escalas mediante el estudio del comportamiento diferencial de reactivos

La equidad es una característica deseable en los instrumentos de medición cuando se realiza evaluación a gran escala. El análisis de comportamiento diferencial de reactivos (DIF, por sus siglas en inglés) sirve para confirmar la equidad de un instrumento de medición. El análisis de DIF se basa en estadísticos asociados con la cantidad de constructo de un reactivo. En preguntas de contexto, un reactivo con DIF se puede dar por problemas del reactivo como la ambigüedad o mala comprensión; o bien, por elementos particulares de la población: culturales, sociales, económicos, etcétera.

A continuación se presenta el mecanismo cuantitativo para el diagnóstico de DIF que está basado en el Modelo de Teoría de respuesta al ítem de crédito parcial con facetas (MCPF). Este modelo permite agregar términos (facetas) que modifican a la localización del reactivo y que indican diferentes factores que pueden tener influencia sobre su localización (Wu, Adams, Wilson y Haldane, 2007).

Definición

De acuerdo con Angoff (1993), un reactivo con DIF muestra diferentes propiedades estadísticas para distintos grupos luego de controlar las diferencias en las localizaciones de los grupos. Por su parte, Wilson (2005) da una definición más intuitiva: un reactivo presenta DIF cuando sujetos en distintos grupos, pero con cantidad de constructo similar, tienden a contestar el reactivo de forma diferente. La idea central es que la cantidad de constructo se deja fija. Sin embargo, el primer autor no hace referencia al análisis cualitativo del reactivo. En la práctica, expertos en los fundamentos teóricos del constructo

analizan los reactivos diagnosticados con DIF. Esto se hace para tomar la decisión de incluir o descartar tal reactivo para la construcción de las escalas.

Grupos

Los grupos comparados se definieron por género (hombres-mujeres); y por población indígena-no indígena, a fin de identificar si los reactivos presentaban sesgo hacia algún grupo en particular.

Estimación

El análisis de DIF se realizó con el programa ACER ConQuest®, de acuerdo con lo que recomienda el Manual de operación en la sección de detección de DIF (ACER Conquest® [Computer Software], 2007). El MCPF modela la probabilidad en categorías adyacentes de que el alumno n en el grupo m con localización en la escala del rasgo latente θ_n conteste la opción k del reactivo i en lugar de la opción $k-1$:

$$\log_e\left(\frac{p_{nimk}}{p_{nim(k-1)}}\right) = \theta_n - (\delta_i - \rho_m - \gamma_m - \gamma_k) \quad \text{con } k = 1, \dots, m_i$$

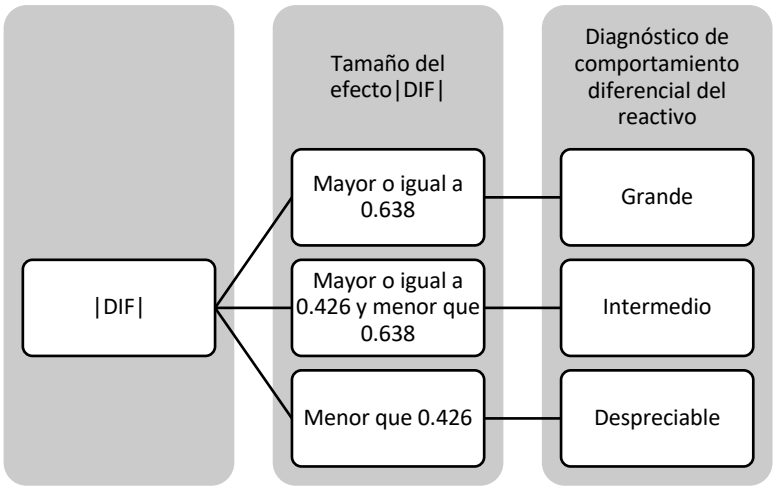
Donde:
p_{nimk} es la probabilidad de que el alumno n tenga el puntaje k en el grupo m del reactivo i,
$p_{nim(k-1)}$ es la probabilidad de que el alumno n tenga el puntaje k-1 en el grupo m del reactivo i,
m_i es el número de opciones del reactivo i.

El MCPF tiene cuatro parámetros que interactúan con el parámetro de localización del alumno. Dos términos de efectos principales: reactivo δ_i y grupo ρ_m, los cuales estiman la localización para cada reactivo y la media para cada grupo. Un término de interacción: reactivo*grupo γ_m, que modela la variación en la localización del reactivo entre los grupos. El cuarto término: reactivo*grupo*step, modela la diferencia debida a cada paso de reactivo y cada grupo γ_k.

El estadístico para el diagnóstico se etiquetó como |DIF|. Este estadístico es la diferencia aritmética en valor absoluto de los parámetros estimados de interacción reactivo*grupo γ_m o de ubicación media para cada grupo. En particular, para dos grupos, en la parametrización de Conquest® basta con tomar el doble del valor del parámetro estimado.

Criterio para el diagnóstico

Cuando el tamaño de muestra es grande, los datos tienen suficiente potencia para detectar diferencias muy pequeñas, de modo que las pruebas estadísticas tradicionales frecuentemente muestran significancia. La prueba tradicional consiste en tomar como estadístico de prueba la diferencia entre las localizaciones de los reactivos para los dos grupos (o la diferencia entre los parámetros de interacción reactivo*grupo) y verificar si son más grandes de lo esperado dividiendo esta diferencia por su error estándar, como en una prueba tipo z. De esta manera, todos los reactivos serían diagnosticados con DIF. Por lo anterior, se consideró el tamaño del efecto, que es la diferencia en valor absoluto entre los parámetros estimados para la interacción reactivo*grupo, la cual se denominó: |DIF|. El criterio de clasificación está basado en Wilson (2005) y el Educational Testing Service (An adjustment for Sample Size in DIF Analysis):



Los reactivos con diagnóstico de DIF grande deben tomarse con precaución. El tratamiento para un reactivo diagnosticado con DIF grande puede variar. Se recomienda consultar la opinión de expertos en los fundamentos teóricos del constructo. En ciertos casos, se decidió eliminar el reactivo de la escala. En otros casos, se conservó el reactivo y se reportó para que se tomaran los resultados con cautela.

Cabe resaltar que el análisis de DIF es un procedimiento posterior a la verificación de la bondad de ajuste de los reactivos. Los reactivos con ajuste deficiente pueden dar un diagnóstico espurio.

3.4 Recomendaciones que describen las características psicométricas de los reactivos dentro del modelo de medición de crédito parcial

Varios autores han desarrollado recomendaciones que describen las características psicométricas del reactivo bajo el Modelo de Rash de crédito parcial (MCP). Estas recomendaciones están basadas en los valores que toman las estimaciones de los parámetros del modelo y en las estadísticas numéricas de ajuste del mismo. Además, se pueden explorar las representaciones gráficas de la relación entre la localización del reactivo y del alumno o las curvas características de probabilidad, de probabilidad acumulada o de puntaje esperado del modelo. Cabe destacar que las recomendaciones apoyadas en resultados gráficos son limitadas en comparación con las basadas en rangos y criterios cuantitativos.

3.4.1 Recomendaciones paramétricas y de ajuste del modelo

Los lineamientos que se presentan a continuación están basados en las presentaciones que hacen Dimitrov (2012), Wu y Adams (2007) y Wilson (2005), considerando adaptaciones para el MCP.

- **Al menos 10 observaciones en cada categoría**

Debe haber al menos 10 alumnos en cada categoría de los reactivos. Cuando la frecuencia en la categoría es baja, los umbrales se estiman de manera pobre y potencialmente inestable.

- **Distribución regular de las observaciones**

La distribución de las observaciones en las categorías dentro de cada reactivo puede ser de tipo unimodal, que tiene un pico central o con categorías extremas, o bimodal, que tiene picos en las categorías extremas. Por lo tanto, se debe verificar la distribución empírica de las observaciones para cada reactivo en las categorías.

- **Promedio de la localización de los alumnos debe ser monótono con las categorías**

El *promedio de la localización de los alumnos en cada categoría* sirve para identificar si existe un desorden en las categorías, se esperaría que valores altos de categoría tengan *promedio de la localización de los alumnos en cada categoría* alto.

- **La correlación punto biserial es creciente con las categorías**

Según Wu y Adams (2007) se esperaría que a mayor categoría, mayor puntaje total. Esto implica que la correlación punto biserial debería incrementarse conforme se incrementa la categoría. Sin embargo, en Wu, Adams, Wilson y Haldane (2007) se reporta que algunas veces los puntos biseriales no están ordenados cuando una proporción muy pequeña o muy grande de las respuestas al reactivo están en una categoría.

- **Estadísticos de ajuste basados en residuales**

El infit es un estadístico que indica que la pendiente de la curva observada es la misma que la esperada o teórica. Se espera tener valores cercanos a uno. Wilson (2005) plantea como intervalo aceptable para el infit 0.75 a 1.33.

El outfit es un estadístico de ajuste que, de acuerdo con Dimitrov (2012), permite detectar desorden en las categorías de cada reactivo. Esta medida puede indicar el grado con el cual las categorías tienen terminología ambigua o si la cantidad de categorías es excesiva. Cuando el valor del outfit es cercano a 1 se obtiene un buen ajuste y, a lo más, es recomendable que sea igual a 2 (Linacre, 2002).

- **Los umbrales τ´s avanzan monótonamente con las categorías**

La probabilidad de obtener una calificación alta debería incrementarse con el aumento en la localización θ del alumno (Dimitrov, 2012). Por lo cual se debe verificar que los τ´s de las categorías se incrementan conforme aumenta la categoría.

Si los umbrales no tienen avance monótono entre categorías adyacentes se dice que hay desorden de umbrales, desorden de calibraciones de paso o, simplemente, desorden de paso (Linacre, 2002). Las causas pueden ser, según Andrich, de Jong y Sheridan (1997), que no es posible definir un rango modal para alguna categoría, que no se utilicen todas las categorías o por un desorden empírico de las categorías.

- **Amplitud de la categoría**

Cuando los parámetros $\tau_1, \tau_2, \dots, \tau_{m_i-1}$ están ordenados respecto a las m_i categorías, la amplitud de la categoría intermedia k $(k = 2,3, \dots, m_i)$ se define como la distancia entre los umbrales para esa categoría. El intervalo (τ_k, τ_{k-1}) es el intervalo modal de la categoría, esto es, el intervalo en la escala $logit$ donde la categoría k es más probable de ser observada que cualquier otra. Dimitrov (2012) basado en resultados de Linacre (2002), sugiere que una amplitud adecuada es de 1.4 a 5.0 $logits$. Linacre (2002) notó que "cuando la distancia es mayor a 5.0 $logits$ la información dada en el centro del reactivo

es menor que la mitad de lo que daría una simple dicotomía". Por otro lado, un intervalo modal angosto (menor a $1.4\ logits$) puede indicar que la categoría representa un segmento muy angosto de la variable latente o corresponde a un concepto que está definido de manera deficiente en la mente de los que responden.

- **Qué hacer cuando no se cumplen algunas recomendaciones**

Si los umbrales $\tau's$ no están ordenados o si otros estadísticos de ajuste muestran valores fuera de rango en los datos y el modelo, entonces puede ser que colapsar categorías revele el número efectivo y el ordenamiento de categorías (Wright, 1994). Esto se puede hacer mediante un estudio de seguimiento usando el nuevo formato de categorías.

3.4.2 Recomendaciones gráficas del modelo

La literatura hace referencia a gráficos que son de utilidad para la interpretación del modelo, sin embargo es limitada la información que se extrae del análisis visual de dichos gráficos (Dimitrov, 2012). Es por esta razón que a continuación se plantean dos notas a observar, tomadas de Wu y Adams (2007) y Bond y Fox (2001). En estas notas se revisa el ajuste del modelo y la orientación de los resultados de los alumnos de forma visual.

Curvas características de probabilidad acumulada de los reactivos. Son útiles para observar el comportamiento de los reactivos, en ellas se grafica por cada categoría la probabilidad de obtener dicha categoría o una mayor dependiendo de la localización (parámetro θ del modelo). En la misma gráfica se muestran dos curvas: la dada por el modelo ajustado (curva teórica) y la del modelo de predicción (curva empírica). Wu y Adams (2007) mencionan que ambas curvas características del reactivo deben ser cercanas.

Comparación de las distribuciones de alumnos y reactivos (Mapa de Wright). Estos valores se muestran gráficamente en el mapa de distribución latente o Mapa de Wright, del lado izquierdo las localizaciones estimadas de los alumnos y del lado derecho los umbrales de Thurstone de los reactivos.

El umbral de Thurstone para el reactivo i y la categoría k es la localización sobre la escala del rasgo latente en la cual la probabilidad de alcanzar la categoría k o una mayor alcanza 0.5 (Wu y Adams, 2007). Debido a que es una medida que acumula probabilidades, estos umbrales siempre están ordenados respecto a las categorías del reactivo y permiten dar una mejor medida de cantidad de constructo de las categorías de los reactivos que los parámetros del MCP δ_{ik} que no necesariamente están ordenados respecto a las categorías.

En la parte superior del lado izquierdo del Mapa de Wright se representan los alumnos con mayor cantidad de constructo y en la parte inferior se encuentran los alumnos con menor cantidad de constructo. Análogamente, en la parte superior del lado derecho están los reactivos con mayor cantidad de constructo y en la parte inferior se encuentran los reactivos con menor cantidad de constructo. Cabe mencionar que el rango del mapa está en escala logit, un rango aceptable es de -3 a +3 (Bond y Fox, 2001).

La medida se considera adecuada cuando hay un traslape entre los rangos de valores para ambas distribuciones (Dimitrov, 2012; Bond y Fox, 2001). En evaluaciones de logro, cuando la distribución de localizaciones de los alumnos se concentra en un rango de valores menores a las localizaciones de los

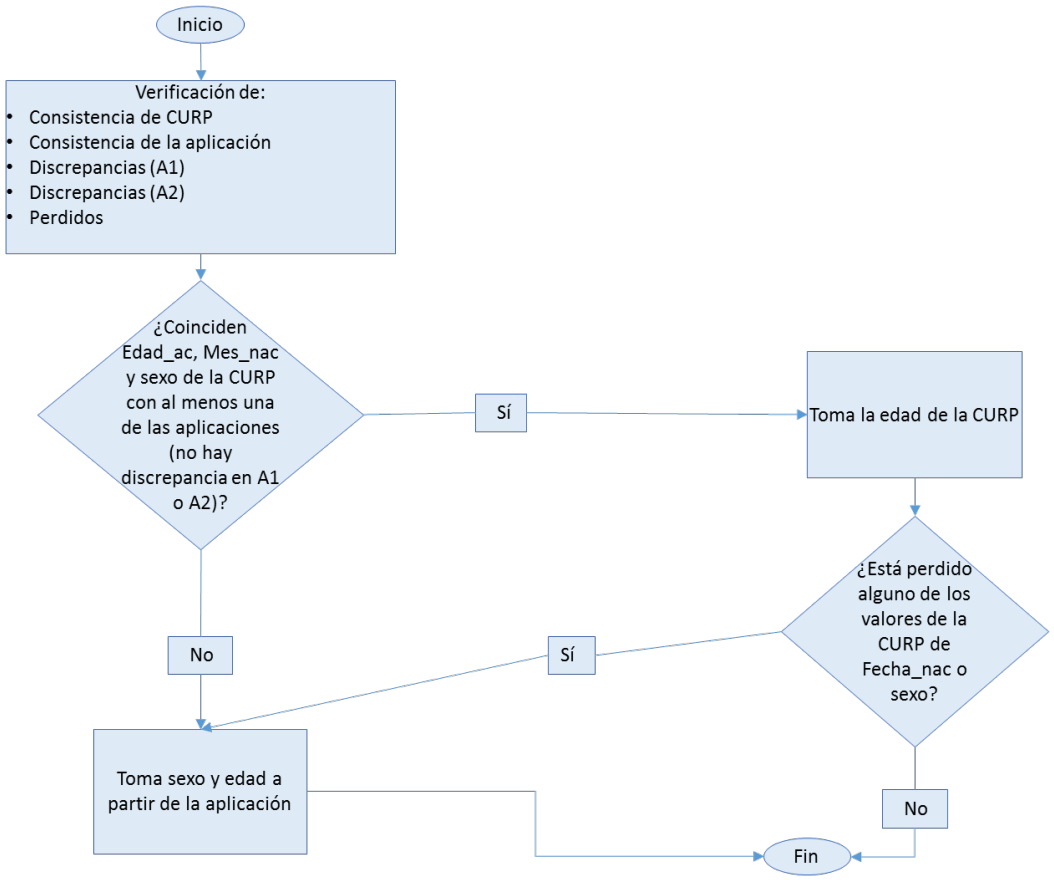
reactivos, se dice que el instrumento es relativamente difícil; en el caso donde la distribución de localizaciones de los alumnos se localiza en un rango de valores mayores a las localizaciones de los reactivos, el instrumento es relativamente fácil (Bond y Fox, 2001). En evaluaciones de contexto diríamos que la cantidad de constructo del instrumento es mayor a la de la población en el primer caso o que la cantidad de constructo en la población es mayor a la del instrumento en el segundo caso.

Apéndice C1: Construcción de las variables de edad y sexo en Planea EMS, 2017

Proceso de construcción de las variables Edad y Sexo para PLANEA EMS 2017

A continuación se describe un algoritmo para construir las variables Edad y Sexo en PLANEA EMS 2017. En el Diagrama 1 se presenta el proceso para construir la variable Edad.

Diagrama 1. Flujo del proceso de construcción de la variable Edad.



Se dio prioridad a valores no discrepantes. Si las variables, basadas en la CURP, Edad_AC, Mes_nac y Sexo coinciden al menos con una de las aplicaciones, entonces se toman la edad y el sexo de la CURP por ser las más precisas.

Luego, si las variables Edad_AC, Mes_nac y Sexo no coinciden en al menos una aplicación y en la CURP, se toman la edad y el sexo con base en la aplicación, utilizando el algoritmo de PLANEA 2015. Si hay errores de consistencia en la aplicación y no coinciden los valores de la CURP con al menos una de las aplicaciones, se tienen tres valores de edad y sexo diferentes; en tal caso, se optó por tomar el valor observado en la aplicación ya que en el momento de la aplicación el examen dirigido a un alumno se le pudo haber dado a otro.

Por último, en caso de que algún valor estuviera perdido en la CURP, se usó la aplicación y, si hay al menos un valor válido en la aplicación (de edad o sexo), los datos se consideraron.

Es importante mencionar que sólo se usó una fuente de información, ya sea la CURP o la aplicación. Tanto la edad como el sexo se tomaron sólo de la CURP o de la misma aplicación.

Construcción de las variables

Para la construcción de las variables se utilizó un proceso de doble validación.

Una parte de la validación se realiza dentro de la aplicación pues se tienen dos fuentes de información: el cuestionario de Lenguaje y Comunicación y el de Matemáticas. A la coincidencia entre los valores de Edad_AC, Mes_nac y Sexo en las dos aplicaciones, se le denominó consistencia dentro de la aplicación. Otra parte consideró la consistencia de la CURP. Se definieron los siguientes errores de la CURP:

1. La longitud de la cadena es diferente de 18 caracteres.
2. Algún caracter en la cadena no es del tipo que debería ser (letra o número). Posición, por ejemplo: dígito, vocal, consonante.
3. Los dígitos en la posición 5-10 no representan una fecha válida.
4. El penúltimo caracter no es numérico para fechas previas a 2000.
5. El código de verificación es incorrecto.

Estos errores se codificaron en un programa que en caso de que ocurran varios errores de los mencionados, envía el menor valor de los códigos asignados.

Posteriormente a esta doble validación se definió un criterio de Discrepancia. Esta discrepancia fue entre los valores de la CURP que se compararon con los de Edad_AC, Mes_nac y Sexo obtenidos en la aplicación. Se comparó la información de la CURP contra la aplicación de L y C (DiscripanciasA1) y contra la aplicación de Matemáticas (DiscrepanciasA2).

En caso de que la CURP no presentara errores de consistencia con respecto a los criterios del 1 al 5, se utilizó la fecha de nacimiento de la CURP para estimar la edad del alumno, siempre que no hubiera discrepancia del tipo A1 o A2.

En caso de que la CURP presentara algún error de validez (del 1 al 5) se tomaron las siguientes decisiones para estimar la edad como se presentan en la Tabla 1.

Tabla 1. Propuesta de decisión para asignación de la variable Edad en PLANEA EMS 2017

Error en CURP	Consistencia aplicación (errores)	Discrepancias A1	Discrepancias A2	Perdidos (aplic)	Frecuencia	Decisión
4-5	no	no	no	no	1 285	Usa CURP
4-5	no	sí	sí	no	218	Usa Aplicación
4-5	no	sí	sí	sí	14	Usa Aplicación
4-5	sí	no	sí	no	22	Usa CURP
4-5	sí	no	sí	sí	162	Usa CURP
4-5	sí	sí	no	no	15	Usa CURP
4-5	sí	sí	no	sí	36	Usa CURP
4-5	sí	sí	sí	no	7	Usa Aplicación
4-5	sí	sí	sí	sí	53	Usa Aplicación
1-3	no	no	no	no	716	Usa CURP
1-3	no	no	no	sí	10	Usa CURP
1-3	no	sí	sí	no	3 264	Usa Aplicación
1-3	no	sí	sí	sí	20	Usa Aplicación
1-3	sí	no	sí	no	9	Usa CURP
1-3	sí	no	sí	sí	142	Usa CURP
1-3	sí	sí	no	no	6	Usa CURP
1-3	sí	sí	no	sí	398	Usa CURP
1-3	sí	sí	sí	no	121	Usa Aplicación
1-3	sí	sí	sí	sí	234	Usa Aplicación
0	no	no	no	no	91 670	Usa CURP
0	no	no	no	sí	42	Usa CURP
0	no	no	sí	no	6	Usa CURP
0	no	sí	no	no	1	Usa CURP
0	no	sí	sí	no	2 944	Usa CURP
0	no	sí	sí	sí	777	Usa Aplicación
0	sí	no	sí	no	934	Usa CURP
0	sí	no	sí	sí	10 008	Usa CURP
0	sí	sí	no	no	660	Usa CURP
0	sí	sí	no	sí	2 835	Usa CURP
0	sí	sí	sí	no	154	Usa Aplicación
0	sí	sí	sí	sí	937	Usa Aplicación

Nota: “No” significa que no hay error. “Sí” significa que, al menos, hay un error.

Terminología

Error en CURP: Se obtiene un número del 0 al 5 que representa un error, se usó el algoritmo proporcionado por la Dirección de Diseño de Productos del Sistema. Se tomaron tres categorías: (0) sin error, errores graves (1-3) y no tan graves (4-5).

Consistencia aplicación (errores): Se consideró la coincidencia entre las observaciones de Edad_AC, Mes_nac y Sexo para las aplicaciones de L y C y Mat, en caso de que alguna de estas tres variables no coincidiera en L y C y Mat se determinó que sí hay un error.

Discrepancias A1 y A2: Se comparó la información obtenida en Edad_AC, Mes_nac y Sexo contra la información de la CURP. Como son dos aplicaciones, A1 corresponde a información de la CURP contra L y C, mientras que A2 corresponde a información de la CUPR contra Matemáticas.

Perdidos (en aplicación): Se determinó "sí" cuando hay al menos un valor perdido en la aplicación, ya sea en L y C o en Matemáticas.

Decisión: Determina la fuente de información que se toma para estimar edad y sexo en cada registro. Se le dio prioridad a la no discrepancia. Es decir, la CURP puede tener información válida, pero no corresponder al individuo que contesta la aplicación. Por lo tanto, se considera la información de la aplicación. Dicho de otro modo, si hay un "sí" en Discrepancias A1 y A2, la CURP no coincide con el alumno y por lo tanto no se debe utilizar.

Valores atípicos para la edad

Luego de estimar la edad mediante el proceso descrito, se hizo una última corrección para valores muy pequeños o muy grandes para la edad que se consideraron poco factibles. En caso de que la edad fuera menor a 15 años o mayor a 50, se sustituyó la edad calculada (mediante la CURP) por la edad calculada a partir de la aplicación.

A continuación, en la Tabla 2 se presentan descriptivos de las decisiones que se tomaron para asignar la edad. Casi todas las observaciones se tomaron a partir de la CURP.

Tabla 2. Descriptivo por tipo de fuente informativa para calcular la edad

Decisión	n	Porcentaje
Usa aplicación	8 743	7%
Usa CURP	108 957	93%
Total	117 700	100%

Comparación con datos del análisis en paralelo

El análisis en paralelo sigue el mismo razonamiento de extracción de Fecha de nacimiento y Sexo sobre una variable corregida de CURP, denominada CorrectedCURP, aportado por la Dirección de Diseño de Productos del Sistema. Sin embargo, se detectaron las diferencias que se muestran en la Tabla 3.

Tabla 3. Comparación de valores estimados para Fecha de nacimiento por el método de procedimientos paralelos

Identificador de igualdad de Fechas de Nacimiento	Decisión	n
1	Usa Aplicación	3 188
1	Usa CURP	108 484
0	Usa Aplicación	5 555
0	Usa CURP	473

En total, 111 672 Fecha de Nacimiento coinciden y 6 028 no. Respecto a las observaciones que tienen fechas de nacimiento diferentes estimadas, en la mayoría (5 555 observaciones) no se utilizó la CURP para estimar el sexo y la edad.

En el caso de la asignación de Sexo no hay coincidencia total, lo cual se indica en la tabla 4. Cabe mencionar que en la SAD se consideraron las mismas decisiones que se tomaron para Edad.

Tabla 4. Comparación entre valores estimados para Sexo tomando en cuenta la decisión

Sexo en datos procedimiento paralelo	Sexo en datos SAD	Decisión	Frecuencia	Coincidencia
1	1	Usa Aplicación	3 697	Coinciden los valores
1	1	Usa CURP	51 280	
2	2	Usa Aplicación	3 782	
2	2	Usa CURP	57 653	
9	9	Usa Aplicación	12	
9	9	Usa CURP	20	
1	2	Usa Aplicación	138	No coinciden los valores
1	9	Usa Aplicación	479	
1	9	Usa CURP	3	
2	1	Usa Aplicación	136	
2	9	Usa Aplicación	499	
2	9	Usa CURP	1	

En total, 116 444 estimaciones de Sexo coinciden y 1 256 no.

Terminología

Sexo en datos procedimiento en paralelo: Este procedimiento considera principalmente la CURP como fuente de información, fue realizado por la Dirección de Diseño de Productos del Sistema.
Sexo en datos SAD: Este procedimiento de asignación de la variable Sexo se basa en el algoritmo descrito en este documento, desarrollado en la Subdirección de Análisis de Datos (SAD).
Decisión: Determina la fuente de información que se toma para estimar edad y sexo en cada registro.
Frecuencia: Número de observaciones con la característica descrita mediante las columnas de la tabla.

Construcción de la variable Edad_AT

Se plantea el uso del término "Edad arriba de la típica" denominado en la base pública como Edad_AT, para edades que están dos años o más arriba de la edad que tendría el alumno en el último grado de EMS si ingresara con edad mínima conforme a la ley vigente.

La variable Edad_AT es dicotómica. Vale 1 para alumnos cuya edad que supera por dos años la edad que tendría un alumno si hubiera entrado con la mínima reglamentaria. Esto en la fecha de aplicación y de acuerdo con la duración de su plan de estudios. A continuación se presenta una estimación del porcentaje de alumnos en esta categoría. Esta sección se agregó con el fin de hacer una aproximación, sin ponderar.

Se comparó la variable Edad con una cota definida según la duración de cada programa educativo. Para obtener tal cota se suma la edad del último año de secundaria en la fecha de aplicación (14.33) más la duración del programa de estudios y dos años más. Valores arriba de esta cota se denominaron: Por arriba de la edad típica. También se tomó en consideración que la ley vigente entró en marcha en el ciclo escolar 2006-2007. Los alumnos que ingresaron a primaria previo a esta fecha tenían una legislación que mencionaba 6 años cumplidos a más tardar el 31 de agosto del año en el que iniciaban el ciclo escolar. En la Tabla 5 se presentan los descriptivos por cada una de las cotas.

Tabla 5. Frecuencias de las cotas sin ponderar

Cotas	Número de observaciones	% del total
Perdidos	4 292	3.6%
18.33	4 977	4.2%
19.66	108 429	92.1%
20.66	2	0.002%

Al comparar la edad con la cota establecida por arriba de la edad típica se obtiene la distribución de ésta como se muestra en la
Tabla ***6***.

Tabla 6. Descriptivos por categoría de la variable edad típica

Categorías	Número de observaciones	% del total
Por arriba de la edad típica	6 974	5.93%
Edad típica	106 434	90.43%
Perdidos	4 292	3.65%
Total	117 700	100.0%

En el *Panorama Educativo de México 2016* se presenta un dato de rezago en el tercer grado de EMS de 12.1% en el 2015. Tal valor sería el análogo de los valores Por arriba de la edad típica, en este caso 5.9%.

Edad en años cumplidos por categoría

El índice de Edad en años cumplidos por categoría se denota como EDAD_acc en la base de datos pública y es una recategorización de EDAD_ac que corresponde a las categorías mostradas en la Tabla 7.

Tabla 7. Categorías de la varible edad en años cumplidos

Categoría EDAD_acc	EDAD_ac Educación Media Superior
1	16 o menos
2	17
3	18
4	19 o más

En caso de que el valor de EDAD_ac fuera omitido, también se omitió el valor de EDAD_acc.

Comentarios finales

Este documento reporta el proceso de construcción de las variables Edad y Sexo en PLANEA EMS 2017 mediante el uso de una validación que considera no sólo la CURP sino también las observaciones de las aplicaciones. Se dio preferencia a la coincidencia (no discrepancia) entre valores de la CURP y de la aplicación. En casos de discrepancia se utilizó la información de la aplicación. También se utilizó la información de la aplicación cuando no hubo consistencia en las observaciones dado que en ocasiones las pruebas que van dirigidas a un estudiante, se le dan a otro.

Apéndice C2: Reporte del modelo de medición para la escala de Capital económico. Modelo de crédito parcial y análisis DIF PLANEA EMS 2017

Resultados

Esta escala presenta condiciones favorables para construirla, pero en cuanto a la cantidad de descriptores que puede tenerse en cada nivel que se considere para categorizarla, la información es limitada. Es por ello, que no es recomendable contrastar estimaciones de logro que sirvan para explicar algún aspecto de la marginación, sobre todo si el índice de marginación se concentra en caracterizar a las poblaciones más desfavorables.

Resumen

La escala Capital económico (CE) se construyó utilizando siete reactivos del Cuestionario de estudiantes aplicado en PLANEA EMS 2017 (MA_114 al MA_120). Los reactivos MA_115 y MA_119 se unieron para construir el índice Televisor (TV), en este caso considera cuatro categorías en lugar de tres, lo cual se explica más adelante. En un análisis preliminar de la construcción del MCP, los

reactivos MA_114 y MA_116 mostraron que las categorías 2 y 3 presentaban poca distancia entre ellas, por lo cual se optó por juntar dichas categorías. En el anexo se presentan los resultados de las estimaciones previas a la construcción del modelo de medición: matriz de correlación, valor de alfa de Cronbach, índice de discriminación, proporción de respuesta y valor de pplus. Estos resultados son satisfactorios para todos los reactivos y en algunos casos presentan mejoras con el análisis reportado en el documento "Construcción de escalas PLANEA EMS 2017, etapas 1 y 2".

La medida de la escala se construyó mediante el Modelo de Rasch de crédito parcial (MCP) y se evaluó con el comportamiento diferencial de reactivos (DIF, por sus siglas en inglés) mediante el Modelo de facetas del modelo de crédito parcial (MCPF) respecto a las variables Sexo y Condición indígena.

Se consideraron los estadísticos de ajuste infit, ancho de categoría, correlación punto biserial y valores promedio por categoría para llegar a tal conclusión. Respecto al DIF, todos los reactivos tienen magnitud de DIF despreciable excepto MA_120 respecto a condición indígena de magnitud intermedia.

Reactivos que conforman la Escala

Los reactivos que conforman esta escala son MA_114 a MA_120. El reactivo TV es, en realidad, un índice que se formó a partir de la combinación de los reactivos MA_115 y MA_119. Los reactivos MA_114, MA_116 se codificaron para iniciar en 0 y tienen 4 categorías. Los reactivos MA_117, MA_118 y MA_120 son dicotómicos y se codificaron para las categorías 0 y 1. Estos tres reactivos estaban en sentido inverso en el cuestionario (Tabla 1).

Tabla 1.Reactivos que componen la escala Capital económico

Reactivo	Reactivo en el cuestionario	Etiqueta
TV	Unión de Ma_115 y Ma_119	Índice Televisor (número de aparatos, considerando servicio de cable)
MA_114	Ma_114	Del siguiente listado de bienes, señala cuántos hay en tu casa. Computadora
MA_116	Ma_116	Del siguiente listado de bienes, señala cuántos hay en tu casa. Automóvil
MA_117	Ma_117	En tu casa, ¿cuentas con los siguientes bienes y servicios? Teléfono convencional (fijo)
MA_118	Ma_118	En tu casa, ¿cuentas con los siguientes bienes y servicios? Horno de microondas
MA_120	Ma_120	En tu casa, ¿cuentas con los siguientes bienes y servicios? Internet

La Tabla 2 presenta la codificación de los reactivos, el incremento en el valor de la categoría representa mayor cantidad de constructo.

Tabla 2. Categorías de las variables que componen la escala Capital económico

Variables	Categorías
TV construida a partir de: Ma_115 y Ma_119	0: No tiene televisor, 1: Tiene al menos un televisor (sin servicio de cable o vía satélite); 2: Tiene 1 o 2 televisores y servicio de cable o vía satélite; 3: Tiene 3 o más televisores y servicio de cable o vía satélite
MA_114, MA_116	0: Ninguno, 1: 1, 2: 2 o 3, 3: 4 o más
MA _117, MA _118, MA_120	0: No, 1: Sí

El índice Televisor se codificó como se muestra en la Tabla 3.

Tabla 3. Construcción del índice TV

<table>
<tr><th colspan="2" rowspan="2">(TV) Índice Televisor (número de aparatos, considerando servicio de cable)</th><th colspan="4">MA_119 En tu casa, ¿cuentas con los siguientes bienes y servicios? Televisión por cable o vía satélite</th></tr>
<tr><th>(1) Sí</th><th>(2) No</th><th>(8) Respuesta múltiple</th><th>(9) Respuesta omitida</th></tr>
<tr><td rowspan="5">MA_115 Del siguiente listado de bienes, señala cuántos hay en tu casa. Televisor</td><td>(1) Ninguno</td><td colspan="4">(0) No tiene televisión</td></tr>
<tr><td>(2) 1</td><td rowspan="2">(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3" rowspan="2">(1) Tiene al menos un televisor</td></tr>
<tr><td>(3) 2</td></tr>
<tr><td>(4) 3</td><td rowspan="2">(3) Tiene 3 o más televisores y servicio de cable o vía satélite</td></tr>
<tr><td>(5) 4 o más</td></tr>
<tr><td>(8) Respuesta múltiple</td><td>(2) Tiene 1 o 2 televisores y servicio de cable o vía satélite</td><td colspan="3">( ) Respuesta omitida</td></tr>
</table>

Resultados del análisis previo al MCP

A continuación se presentan los resultados del análisis previo a la construcción del modelo de medición, incorporando las categorías actuales de los reactivos. La Tabla 4 presenta los valores de frecuencia, la proporción de repuesta por categoría y el valor pplus para cada reactivo. En todos los reactivos se presentan valores aceptables, sólo en dos categorías se presenta una proporción de repuesta menor a 0.40.

Tabla 4. Frecuencias, proporción de respuesta y pplus de los reactivos que componen la escala CE

Pregunta		Categoria		Frecuencia	Proporción de respuesta	p plus
MA_114	Del siguiente listado de bienes, señala cuántos hay en tu casa. Computadora	0	Ninguno	26439	0.228	0.379
		1	1	52567	0.454	
		2	2 o 3	31520	0.272	
		3	4 o más	5379	0.046	
		9	Perdido por el sistema	103		
TV	Índice de televisor (número de aparatos, considerando servicio de cable)	0	No tiene televisor	4287	0.037	0.653
		1	Tiene al menos un televisor (sin servicio de cable o vía satélite)	29618	0.256	
		2	Tiene 1 o 2 televisores y servicio de cable o vía satélite	48642	0.420	
		3	Tiene 3 o más televisores y servicio de cable o vía satélite	33346	0.288	
		9	Perdido por el sistema	115		
MA_116	Del siguiente listado de bienes, señala cuántos hay en tu casa. Automóvil	0	Ninguno	41691	0.360	0.318
		1	1	42202	0.364	
		2	2 o 3	27576	0.238	
		3	4 o más	4322	0.037	
		9	Perdido por el sistema	217		
MA_117	En tu casa, ¿cuentas con los siguientes bienes y servicios? Teléfono convencional (fijo)	0	No	47575	0.411	0.589
		1	Sí	68254	0.589	
		9	Perdido por el sistema	179		
MA_118	En tu casa, ¿cuentas con los siguientes bienes y servicios? Horno de microondas	0	No	40663	0.351	0.649
		1	Sí	75162	0.649	
		9	Perdido por el sistema	183		
MA_120	En tu casa, ¿cuentas con los siguientes bienes y servicios? Internet	0	No	35090	0.303	0.697
		1	Sí	80708	0.697	
		9	Perdido por el sistema	210		

La matriz de correlación de Pearson presenta valores por arriba de 0.20. Situación que mejoró con respecto al análisis inicial entregado previamente, pues la correlación entre los reactivos MA_117 y TV era menor a 0.20 (figura 1).

Figura 1. Correlaciones entre los reactivos que componen la escala CE.

Pregunta	MA_114				
MA_114	1.000	TV			
TV	0.355	1.000	MA_116		
MA_116	0.467	0.381	1.000	MA_117	
MA_117	0.320	0.226	0.273	1.000	MA_118
MA_118	0.309	0.304	0.321	0.253	1.000
MA_120	0.440	0.301	0.311	0.465	0.331

El valor de Alfa de Cronbach mejoró respecto al análisis inicial entregado previamente, pues el valor cambió de 0.694 a 0.734. El índice de discriminación se encuentra entre los valores de 0.30 y 0.70 para todos los reactivos (ver Tabla 5).

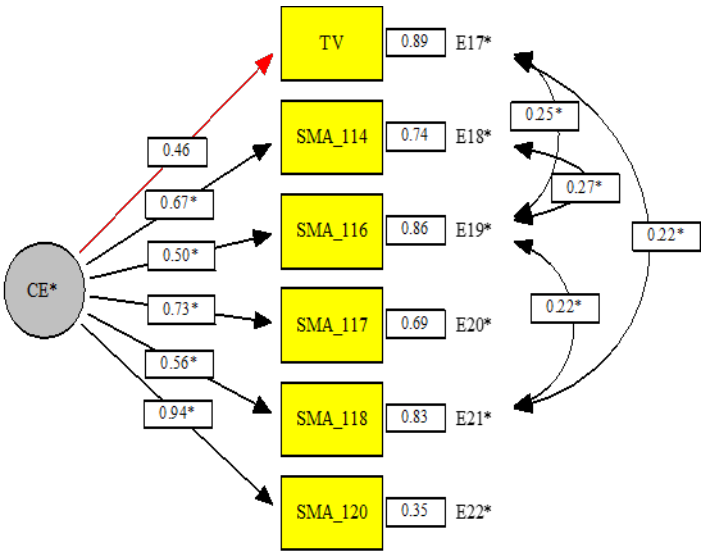
Tabla 5. Análisis de confiabilidad de la escala CE

Pregunta		Alfa de Cronbach	Alfa eliminando el reactivo	Índice de discriminación
MA_114	Del siguiente listado de bienes, señala cuántos hay en tu casa. Computadora	0.734	0.667	0.566
TV	Índice de televisor (número de aparatos, considerando servicio de cable)		0.705	0.463
MA_116	Del siguiente listado de bienes, señala cuántos hay en tu casa. Automóvil		0.681	0.534
MA_117	En tu casa, ¿cuentas con los siguientes bienes y servicios? Teléfono convencional (fijo)		0.714	0.418
MA_118	En tu casa, ¿cuentas con los siguientes bienes y servicios? Horno de microondas		0.712	0.435
MA_120	En tu casa, ¿cuentas con los siguientes bienes y servicios? Internet		0.696	0.526

Respecto al Modelo factorial confirmatorio, éste ajusta con un RMSEA de 0.05, presenta cargas factoriales mayores a 0.30 y los valores de correlación entre errores son del orden de 0.20 (Figura 2).

Figura 2. Modelo factorial confirmatorio de la escala Capital económico.

CFI: 0.994 y RMSEA: 0.05.



Análisis del Comportamiento diferencial de reactivos (DIF)

Se dice que un reactivo no presenta comportamiento diferencial cuando en diferentes grupos de interés, el comportamiento del reactivo es razonablemente similar para sujetos con la misma cantidad de constructo. De lo contrario, decimos que el reactivo tiene DIF (Wilson, 2005).

Base para el diagnóstico

El DIF se diagnosticó con base en Wilson (2005) y el *Educational Testing Service (An adjustment for Sample Size in DIF Analysis*). Se utilizó el programa ConQuest 4 (Adams, Wu y Wilson, 2015). Detalles sobre el ajuste se pueden consultar en el manual del programa. Se calculó la magnitud de las diferencias de los parámetros de interacción reactivo-grupo del Modelo de facetas de tipo Rasch de crédito parcial. Esta magnitud se denominó |DIF|. Estos parámetros aumentan o disminuyen la dificultad del reactivo de acuerdo con el grupo a que pertenecen. De acuerdo con el signo de estos parámetros el DIF puede favorecer a alguno de los grupos comparados. El diagnóstico se realizó utilizando los criterios de la Figura 3. ***Criterios para determinar DIF (Wilson, 2005***

Figura 3. Criterios para determinar DIF (Wilson, 2005).

Valor absoluto de la diferencia del parámetro de interacción reactivo-grupo |DIF|	Mayor o igual a 0.64	Grande
	Entre 0.43 y 0.64	Intermedio
	Menor que 0.43	Despreciable

Diagnóstico entre grupos de población indígena y no indígena. Este análisis corrió con 111 887 observaciones. Todos los reactivos presentan DIF despreciable excepto el reactivo MA_120 que presenta DIF Intermedio (Tabla 6).

Tabla 6. Resultados del DIF por población indígena

ID Conquest	Reactivo	Estimación	DIF en Ubicación Media	Favorece a…	Diagnóstico
1	TV	-0.065	0.13	Indígena	Despreciable
2	MA_114	-0.142	0.284	Indígena	Despreciable
3	MA_116	-0.095	0.19	Indígena	Despreciable
4	MA_117	-0.067	0.134	Indígena	Despreciable
5	MA_118	0.149	0.298	No_Indíg	Despreciable
6	MA_120	0.220*	0.44	No_Indíg	Intermedio

Nota: un asterisco junto a un parámetro indica que éste está restringido.

Diagnóstico por grupos de Sexo. Este análisis corrió con 116 752 observaciones. Todos los reactivos presentan DIF despreciable (Tabla 7).

Tabla 7. Resultados del DIF por Sexo

ID Conquest	Reactivo	Estimación	|DIF| en Ubicación Media	Favorece a…	Diagnóstico
1	TV	0.102	0.204	Mujer	Despreciable
2	MA_114	-0.009	0.018	Hombre	Despreciable
3	MA_116	-0.055	0.11	Hombre	Despreciable
4	MA_117	-0.033	0.066	Hombre	Despreciable
5	MA_118	-0.043	0.086	Hombre	Despreciable
6	MA_120	0.038*	0.076	Mujer	Despreciable

Nota: un asterisco junto a un parámetro indica que éste está restringido.

El modelo de medición

Los cuestionarios de contexto se escalaron usando modelación de Teoría de respuesta al item (TRI). Se utilizó el Modelo de Rasch de crédito parcial (MCP) descrito por Wu y otros (2007), utilizando el Software ConQuest 4 (Adams, Wu, & Wilson, 2015). De manera similar a lo que se describe en el Reporte Técnico de PISA, Capítulo 16 (OECD, PISA 2006. *Technical Report. Programme for International Student Assessment*, 2009).

Parámetros del Modelo de crédito parcial (MCP)

La Tabla 8 presenta los parámetros del MCP ajustado en ConQuest (Adams, Wu, & Wilson, 2015). Se analizan estos resultados utilizando los criterios de Dimitrov (2012: 208-209).

Tabla 8. Análisis del funcionamiento de la escala por parámetros del MCP

Identificador del reactivo	Dificultad del reactivo	τ_1	τ_2	τ_3	Ancho categoría 2: $\tau_2 - \tau_1$	Ancho categoría 3: $\tau_3 - \tau_2$
TV	-0.935	-2.262	0.161	2.102*	2.423	1.939
MA_114	1.156	-2.301	-0.030	2.330*	2.271	2.36
MA_116	1.556	-1.804	-0.336	2.140*	1.468	2.476
MA_117	-0.270					
MA_118	-0.609					
MA_120	-0.898*					

Nota: un asterisco junto a un parámetro indica que éste está restringido.

Los parámetros τ's son crecientes respecto a la categoría. El ancho de categoría aceptable es entre 1.4 y 5.

Tabla 9. Análisis del funcionamiento de la escala por *infit*, correlación punto biserial y puntaje promedio

Reactivo	Casos	Infit	Correlación punto biserial por categoría				Promedio de puntaje por categoría			
			0	1	2	3	0	1	2	3
1 (TV)	115 995	1.13	-.14	-.27	-.14	.47	-1.40	-.69	.02	1.35
2 (MA_114)	115 905	.98	-.43	-.10	.40	.25	-1.06	.00	1.15	2.13
3 (MA_116)	115 885	1.03	-.44	.01	.38	.23	-.79	.19	1.26	2.21
4 (MA_117)	115 931	1.01	-.42	.42			-.62	.72		
5 (MA_118)	115 926	.98	-.43	.43			-.75	.66		
6 (MA_120)	115 896	.87	-.53	.53			-.97	.66		

El infit es un estadístico que proporciona información sobre qué tanto se ajustan los datos al modelo teórico. Se espera tener valores cercanos a 1, Wilson (2005) plantea dos cotas para el infit: mayor a 0.75 y menor a 1.33; en este análisis se usarán como referencia estas cotas.

Los valores de infit para los reactivos están cerca del valor esperado de 1. Los valores de correlación punto biserial son crecientes respecto a las categorías del reactivo, excepto el reactivo 1 (TV) en la categoría 0 a la 1 que decrece. Por otro lado, el puntaje promedio por categoría es creciente respecto a las categorías para todos los reactivos (tTabla 9).

Gráficos del ajuste del Modelo de crédito parcial (MCP)

En esta sección se presentan las curvas de probabilidad acumulada y las curvas características de cada reactivo en el modelo ajustado. Además, se presenta el Mapa de Wright a 2/3 de probabilidad de acierto.

Curvas de probabilidad acumulada (CC) para los reactivos de la escala CE

A continuación se presenta la CC de cada reactivo y salto de categoría. Todos los reactivos tienen ajuste aceptable con MNSQ o infit dentro del rango de .75 (=3/4) y 1.33 (4/3) utilizado por Wilson (2005).

Gráfico 1. Curva de probabilidad acumulada para el reactivo TV

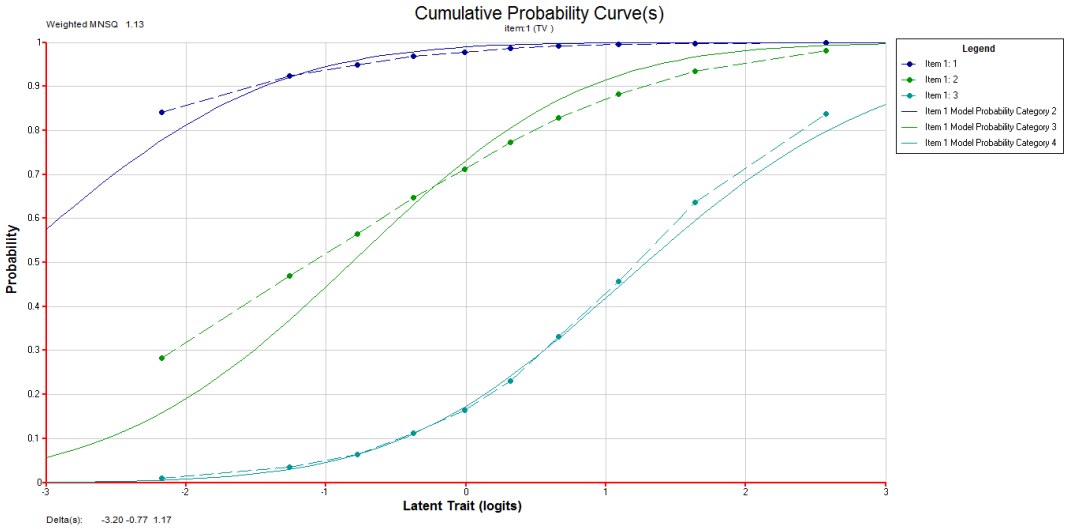


Gráfico 2. Curva de probabilidad acumulada del reactivo MA_116

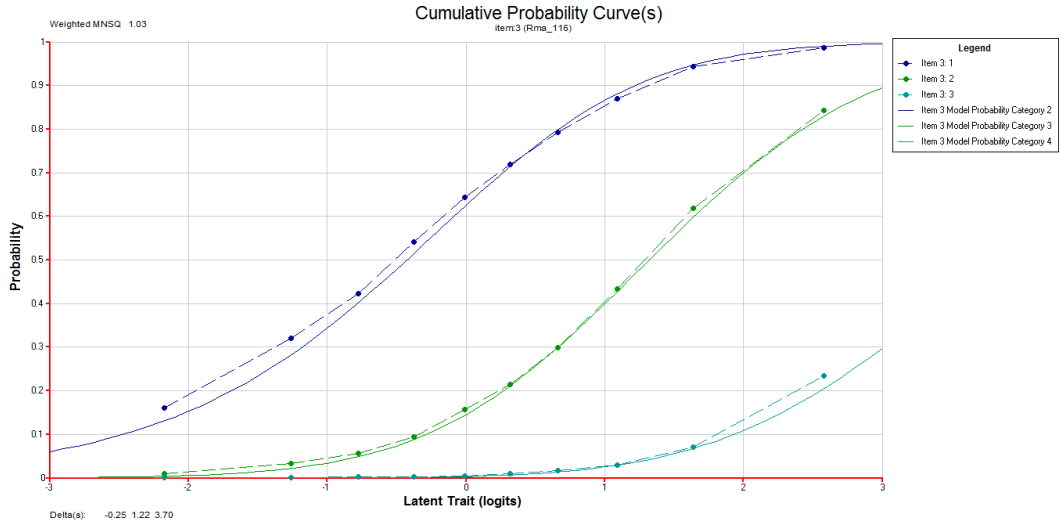


Gráfico 3. Curva de probabilidad acumulada del reactivo MA_114

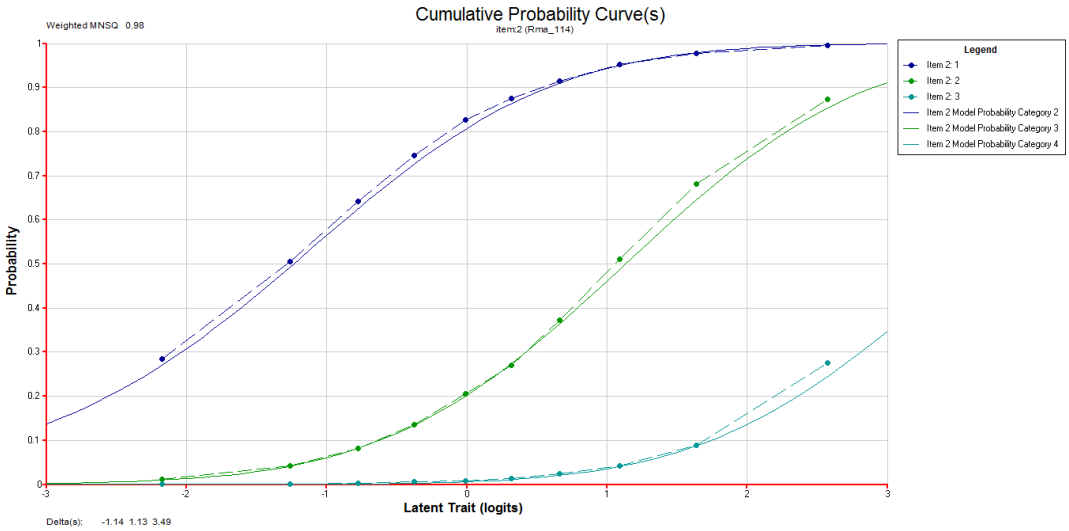


Gráfico 4. Curva de probabilidad acumulada del reactivo MA_117

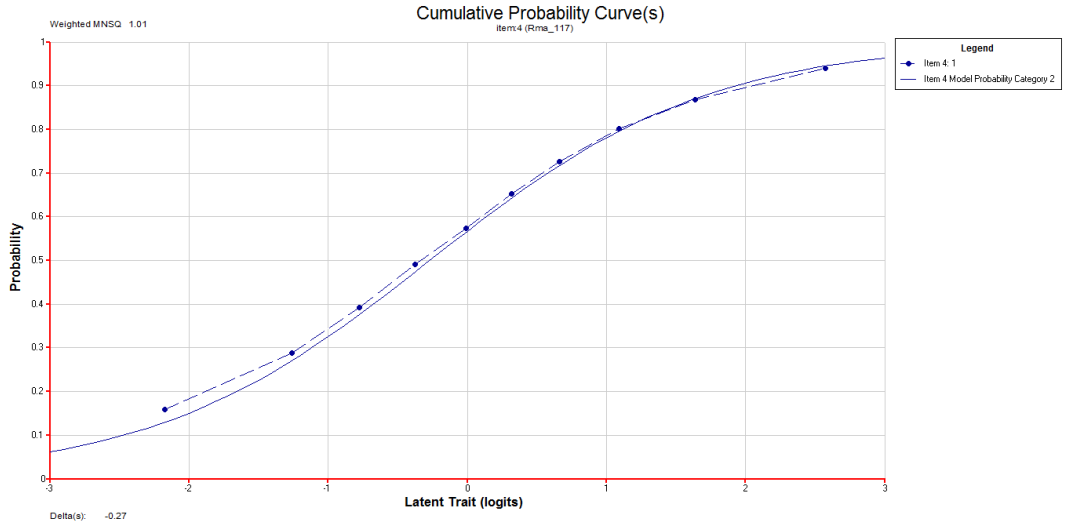


Gráfico 5. Curva de probabilidad acumulada del reactivo MA_115

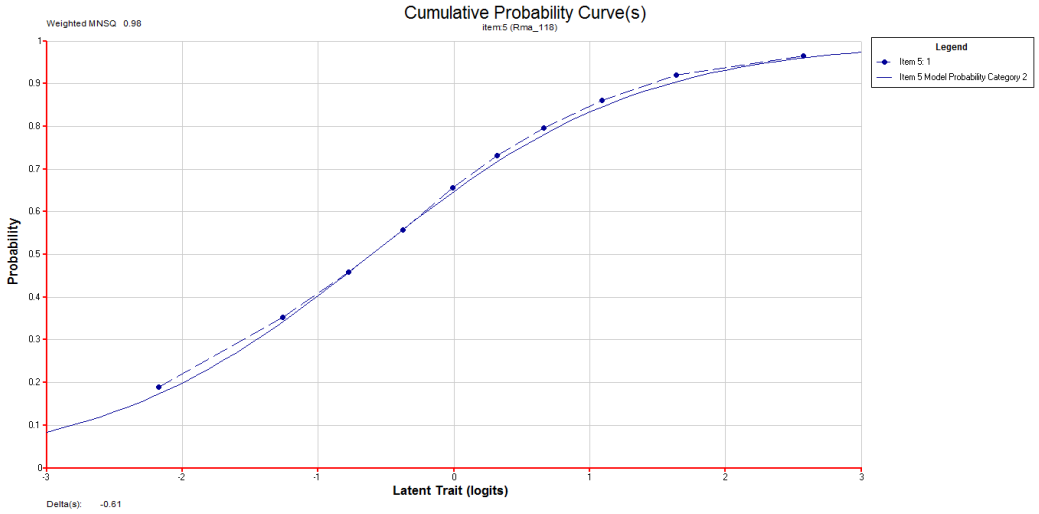
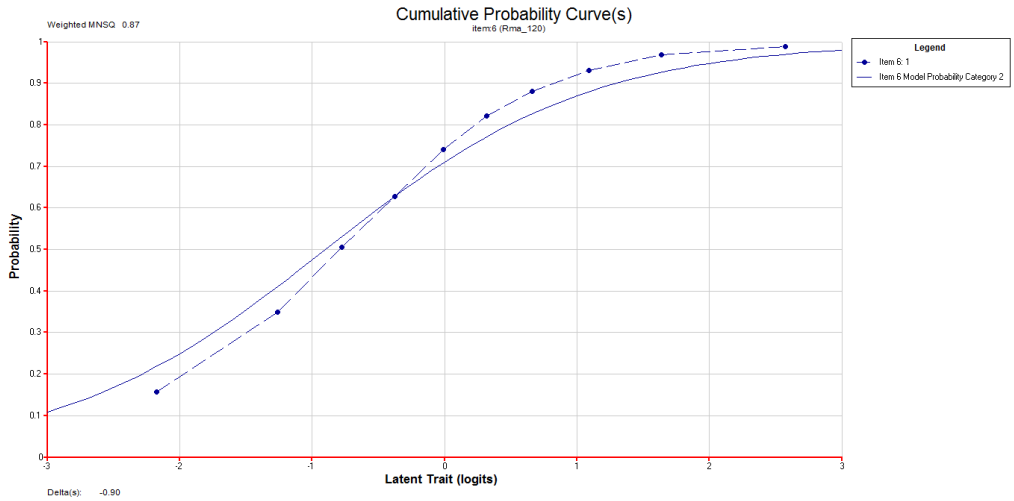


Gráfico 6. Curva de probabilidad acumulada del reactivo MA_120



Curvas características de probabilidad (ICC) para los reactivos de la escala CE

Se presentan las curvas características de probabilidad sólo para los reactivos que tienen más de dos categorías: TV, MA_114 y MA_116. En los reactivos de dos categorías, la curva coincide con la acumulada presentada en la sección anterior. Estas curvas muestran lo que se vio en los estadísticos de infit, el ajuste es aceptable.

Gráfico 7. Curvas características por categoría del reactivo MA_114

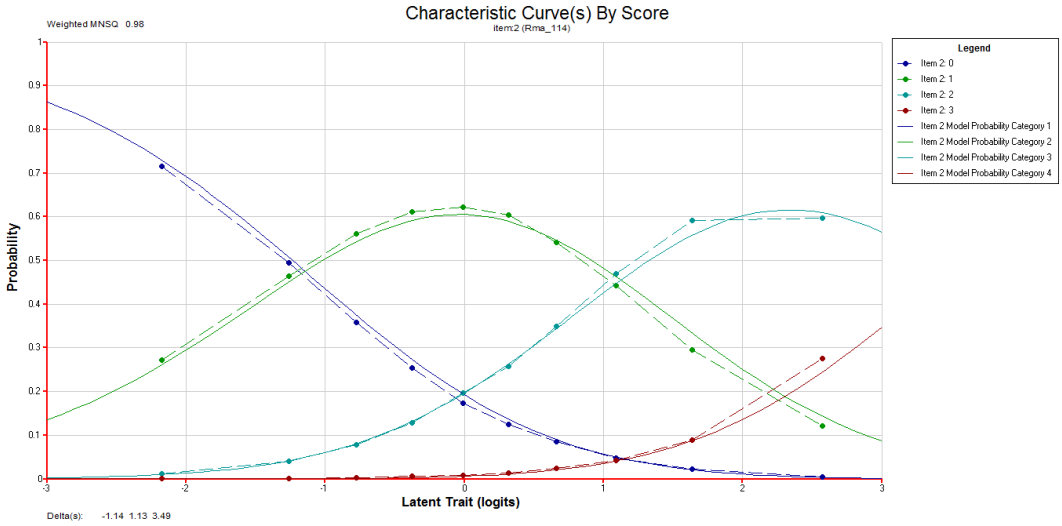


Gráfico 8. Curvas características del reactivo TV por categoría

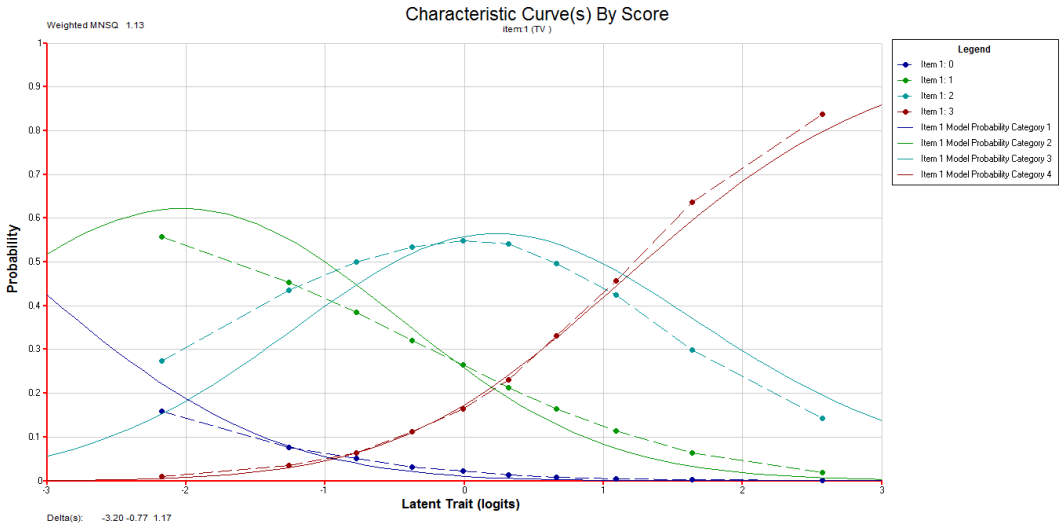
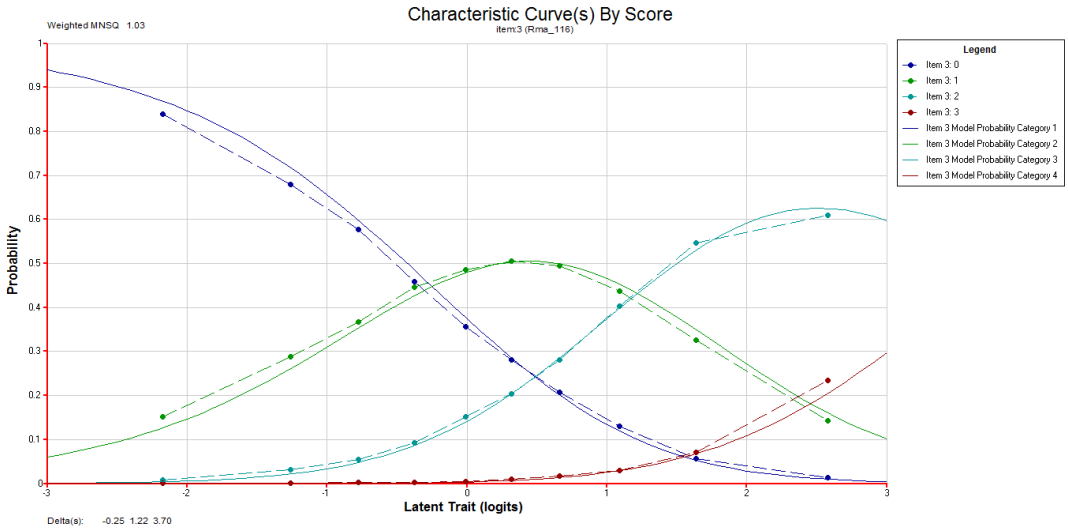


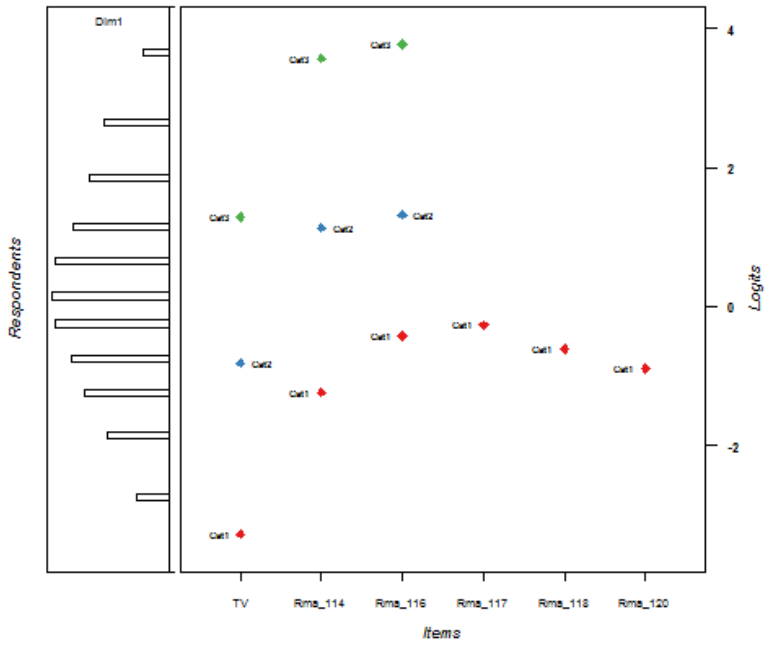
Gráfico 9. Curvas características del reactivo MA_116 por categoría



En esta gráfica se puede ver que la longitud de las categorías es mayor a 1.4.

Mapa de Wright

Gráfico 10. Mapa de Wright de la escala Capital económico



En este gráfico se observan los umbrales al 67% de probabilidad de acierto que se utilizan en la construcción de puntos de corte. Los reactivos y la población están a la misma altura, sobre todo para el reactivo TV. Las categorías de los reactivos MA_114 y MA_116 están más arriba de lo deseable, no discriminan a la población con puntajes bajos, esto puede ser debido a que consideran artículos de “lujo” (número de automóviles y computadoras).

Bibliografía

Adams, R., Wu, M., y Wilson, M. (2015). *Acer ConQuest: Generalised Item Response Modelling Software.* [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.

Andrich, D., de Jong, J., y Sheridan, B. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. En J. Rost, & Langeheine R. (Ed.), *Applications of latent trait and latent class models in the social sciences.* New York, NY: Waxmann Verlag.

Angoff, W. (1993). Perspectives on differential item functioning Methodology. En P. W. Holland, y H. Wainer, *Differential item functioning* (págs. 3-23). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Asparouhov, T., y Muthén, B. (2009). Exploratory Structural Equation Modelling. *Structural Equation Modeling: A Multidiciplinary Journal, 16*(3), 397-438.

Bentler, P., y Wu, E. (2014). EQS 6.2 for Windows (Build 107). Multivariate Software, Inc.

Bock, R. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*(46), 443-459.

Boomsma, A., van Duijn, A. y Snijders, T. (Eds.). (2001). *Essays on Item Response Theory (Lecture Notes in Statistics).* New York: Springer-Verlag.

Bond, T., y Fox, C. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* London: Lawrence Erlbaum Associates, Inc.

Boone, W., Staver, J., y Yale, M. (2014). *Rasch Analysis in the Human Sciences.* Springer.

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (Segunda edición ed.). New York : The Guilford Press.

Cattell, R. (1966). The Screen Test for the Number of Factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Crocker, L. (2008). *Introduction to Classical and Modern Test Theory.* Ohio, USA: Cengage Learning.

Dimitrov, D. (2012). *Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields.* Alexandría: Wiley.

Documento rector del Plan Nacional de Evaluación de los Aprendizajes, PLANEA. http://planea.sep.gob.mx/content/general/docs/2015/PlaneaDocumentoRector.pdf.

Dunn, G. (2004). *Statistical Evaluation of Measurement Errors: Desgin and Analysis of Reliability Studies.* London: Arnold.

ETS. Educational Testing Service. (s.f.). *An adjustment for Sample Size in DIF Analysis*. Recuperado el 19 de Abril de 2016, de http://www.rasch.org/rmt/rmt203e.htm

Everitt, B., Skrondal, A. (2010). *The Cambridge Dictionary of Statistics. Fourth Edition.* Cambridge: Cambridge. University Press.

Guilleux, A. et al. (2014). Power and Sample Size Determination in the Rasch Model: Evaluation of the Robustness of a Numerical Method to Non-Normality of the Latent Trait. *PLOS ONE*, 6.

Hair, J., Black, W., Babin, B. y Anderson, R. (2009). *Mutivariate Data Analysis* (Séptima edición ed.). Pearson.

Harrington, D. (2009). *Confirmatory Factor Analysis.* New York: Oxford, Univesity Press.

Hoyle, R. (1995). *Structural Equation Modeling. Concepts, Issues, and Applications.* SAGE Publications.

Joncas, M., y Foy, P. (2004). TIMSS 2003 Sampling Design. En Martin, M., Mullis, I., y Chrostowski, S. (eds.). *TIMSS 2003 Technical Report* Martin. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Joncas, M. (2007). PIRLS 2006 Sample Design. En Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (eds). *PIRLS 2006 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Joncas, M., y Foy, P. (2012). Sample Design in TIMSS and PIRLS. En Martin, M. y Mullis, I. (eds.). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Jöreskog, K. (1993). Testing structural equation models. En K. A. Bollen y L. JS, *Testing structural equation models* (págs. 294-316). Newbury Park, CA: Sage.

Kish, L. (1965). Survey Sampling. John Wiley & Sons, Inc.

Kaplan, D. (1989). Model Modification in Covariance Structure Analysis: Application of the Expected Parameter Change Statistic. *Multivariate Behavioral Research, 24*(3), 258-305.

Kaplan, D. (2000). *Structural Equation Modeling. Foundations and Extensions, Advanced Quantitative Techniques in Social Sciences.* Sage Publications, Inc.

Kline, R. (2005). *Principles and Practice of Structural Equation Modeling.* The Guilford Press.

LaRoche, S., y Foy, P. (2016). timss 2015 Sample Implementation in timss 2015. En Martin, M., Mullis, I., & Hooper, M. (eds.). Methods and procedures in timss 2015. Chestnut Hill, MA: timss & pirls International Study Center, Boston College.

Little, R., y Rubin, D. (2002). Statistical Analysis with Missing Data, 2nd Edition. Wiley Series in Probability and Statistics.

Lohr, S. (2000). *Muestreo: diseño y análisis*. International Thomson Editores.

Ludlow, L., y O'Leary, M. (1999, August). Scoring omitted and not-reached items: practical data analysis implications. *Educational and Psychological Measurement, 59*(4), 615-629. Recuperado el 9 de mayo de 2019 de: http://epm.sagepub.com/content/59/4/615.full.pdf+html

Lei, P. y Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Instructional Topics in Educational Measurement*, 33-43.

Linacre, J. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Mislevy, R. (1991). Ramdomization-based inferences about latent traits from complex samples. *Psychometrika*, *56* (2), 177-196.

Masters, G. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika Vol. 47, No. 2*, 149 - 174.

Millsap, R. y Olivera-Aguilar, M. (2012). Investigating Measurement Invariance Using Confirmatory Factor Analysis. En R. Hoyle, *Handbook of Structural Equation Modeling* (págs. 380-392). New York: The Guildford Press.

Muthén, L. y Muthén, B. (2015). Mplus. Versión 7.31. Los Ángeles: Muthén & Muthén.

OECD. (2009). *PISA 2006. Technical Report. Programme for International Student Assessment.* Paris: OECD.

OECD. (2014). PISA *2012. Technical Report.* Paris: OECD Publishing.

Raykov, T. y Marcoulides, G. (2011). *Introduction to Psychometric Theory.* New York: Routledge.

StataCorp. (2011). Stata Statistical Software: Release 12. *College Station*. TX: StataCorp LP.

Streiner, D. y Norman, G. (2010). *Health Measurement Scales, 3rd Edition.* Cambridge: Cambridge University Press.

Potter, F. (1990). *Methods for Extreme weights in Sample surveys*. North Carolina: North Carolina State University. Dept. of Statistics.

Richard, J. (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Rubin, D. (1977). *Multiple imputation for nonresponse in surveys.* New York: Wiley Editor.

Rutkowski, L., von Davier, M., y Rutkowski, D. (2014). Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis. Chapman and Hall/CRC.

Särndal, C., Swensson, B., y Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer - Verlag Inc.

Schulz,W., y Sibberns, H. (eds.) (2004). IEA Civic Education Study. Technical Report. The International Association for the Evaluation of Educational Achievement.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis.* Washington, DC: American Psychological Association.

UNESCO-OREALC. (2016). Reporte técnico. Tercer Estudio Regional Comparativo y Explicativo, terce. Santiago de Chile: autor.

Valliant, R., Dever, J., y Kreuter, F. (2013). Practical Tools for Designing and Weighting Survey Samples. New York: Springer.

Van der Linden, W. y Hambleton, R. (1997). *Handbook of modern item response theory.* Springer Science + Business Media, LLC.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach. (Falta editor y lugar*)

Wright, B., y Masters, G. (1982). *Rating scale analysis.* Chicago: MESA Press.

Wright, B., y Stone, M. (1979). *Best test design.* Chicago: MESA Press.

Wright, B., Mead, R., y Draba, R. (1976). *Detecting and correcting test item bias with a logistic response model (Research Memorandum No. 22).* Chicago: MESA Psychometric Laboratory.

Wang, J. y Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus.* Reino Unido: Wiley.

Whittaker, T. (2012). Using the Modification Index and Standardized Expected Parameter Change for Model Modification. *The Journal of Experimental Education, 80*(1), 26-44.

Wilson, M. (2005a). *Constructing Measures. An Item Response Approach.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Wilson, M. (2005b). *Constructing Measures: An Item Response Modeling Approach.* London: Lawrence Erlbaum Associates.

Wright, B. (1977). Solving measurement problems with de Rasch Model. *Journal of Educational Measurement, 14*, 97-116.

Wright, B. (1994). Rasch sensitivity and Thurstone insensitivity to graded responses. *Rasch Measurement Transactions, 8*, 382-383.

Wright, B. y Stone, M. (1979). *Best test design.* Chicago: MESA Press.

Wu, M., Adams, R., Wilson, M. y Haldane, S. (2007). *ACER ConQuest Version 2: Generalised item response modelling software.* Camberwell: Australian Council for Educational Research.

Wu, M. y Adams, R. (2007). *Applying the Rasch Model to Psycho-Social Measurement. A Practical Approach.* Melbourne: Educational Measurement Solutions.

Wu, M., Adams, R. y Haldane, S. (2007). *ACER Conquest® [Computer Software].* Hawthorn, Victoria, Australia: ACER Press. Australian Council of Educational Research. University of California Berkeley.

Instituto Nacional para la
Evaluación de la Educación
México