



# Universidad Autónoma de Baja California

Instituto de Investigación y Desarrollo Educativo

Equidad y análisis de sesgo en las evaluaciones. Los  
Excale de Matemáticas para tercero de secundaria

## TESIS

Que para obtener el grado de

**DOCTOR EN CIENCIAS EDUCATIVAS**

Presenta

Adán Moisés García Medina

Directores de Tesis

Felipe Martínez Rizo

Graciela Cordero Arroyo

Comité Tutorial

Joaquín Caso Niebla

José Felipe Martínez Fernández

Juan Carlos Pérez Morán

Ensenada B. C. México. 24 de febrero de 2017





**Universidad Autónoma de Baja California**  
Instituto de Investigación y Desarrollo Educativo  
Doctorado en Ciencias Educativas



**“Equidad y análisis de sesgo en las evaluaciones.  
Los Excale de Matemáticas para tercero de  
secundaria”**

TESIS

Que para obtener el grado de

**DOCTOR EN CIENCIAS EDUCATIVAS**

Presenta

**Adán Moisés García Medina**

APROBADO POR:

Dr. Felipe Martínez Rizo  
Director

Dra. Graciela Cordero Arroyo  
Co-Directora

Dr. Juan Carlos Pérez Morán

Dr. José Felipe Martínez Fernández

Dr. Joaquín Caso Niebla



Ensenada, B.C., a 19 de diciembre de 2016

**ASUNTO:** Voto aprobatorio al trabajo de tesis para el grado de Doctor en Ciencias Educativas.

**Dra. Alicia Aleli Chaparro Caso López**  
**Coordinadora del Doctorado en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el C. **Adán Moisés García Medina**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo. Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Doctor en Ciencias Educativas, sobre su trabajo titulado:

*“Equidad y análisis de sesgo en las evaluaciones. Los Excale de Matemáticas para tercero de secundaria”.*

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

Una firma manuscrita que parece decir "F. Martínez Rizo".

---

DR. FELIPE MARTÍNEZ RIZO



Ensenada, B.C., a 19 de diciembre de 2016

**ASUNTO:** Voto aprobatorio al trabajo de tesis para el grado de Doctor en Ciencias Educativas.

**Dra. Alicia Aleli Chaparro Caso López**  
**Coordinador(a) del Doctorado en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. Adán Moisés García Medina**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo. Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Doctor en Ciencias Educativas, sobre su trabajo titulado:

*“Equidad y análisis de sesgo en las evaluaciones. Los Excale de Matemáticas para tercero de secundaria”.*

Esperando reciba el presente de conformidad, quedo de Usted.

**Atentamente**

  
**DRA. GRACIELA CORDERO ARROYO**



Ensenada, B.C., a 19 de diciembre de 2016

**ASUNTO:** Voto aprobatorio al trabajo de tesis para el grado de Doctor en Ciencias Educativas.

**Dra. Alicia Alelí Chaparro Caso López**  
**Coordinador(a) del Doctorado en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el C. **Adán Moisés García Medina**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo. Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Doctor en Ciencias Educativas, sobre su trabajo titulado:

*“Equidad y análisis de sesgo en las evaluaciones. Los Excale de Matemáticas para tercero de secundaria”.*

Esperando reciba el presente de conformidad, quedo de Usted.

**Atentamente**



DR. JUAN CARLOS PÉREZ MORÁN



Ensenada, B.C., a 19 de diciembre de 2016

**ASUNTO:** Voto aprobatorio al trabajo de tesis para el grado de Doctor en Ciencias Educativas.

**Dra. Alicia Aleli Chaparro Caso López**  
**Coordinadora del Doctorado en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el C. **Adán Moisés García Medina**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo. Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Doctor en Ciencias Educativas, sobre su trabajo titulado:

*“Equidad y análisis de sesgo en las evaluaciones. Los Excale de Matemáticas para tercero de secundaria”.*

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

DR. JOSÉ FELIPE MARTÍNEZ FERNÁNDEZ



Ensenada, B.C., a 19 de diciembre de 2016

**ASUNTO:** Voto aprobatorio al trabajo de tesis para el grado de Doctor en Ciencias Educativas.

**Dra. Alicia Aleli Chaparro Caso López**  
**Coordinador(a) del Doctorado en Ciencias Educativas**  
**Presente.**

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. Adán Moisés García Medina**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo. Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Doctor en Ciencias Educativas, sobre su trabajo titulado:

*“Equidad y análisis de sesgo en las evaluaciones. Los Excale de Matemáticas para tercero de secundaria”.*

Esperando reciba el presente de conformidad, quedo de Usted.

**Atentamente**

DR. JOAQUÍN CASO NIEBLA

## Agradecimientos

La realización de un trabajo de investigación doctoral es el resultado de una serie de esfuerzos y aportaciones de distintas figuras. Quiero hacer patente mi agradecimiento a:

- Mis directores de tesis, Felipe y Graciela, quienes con toda su calidez humana me brindaron asesoría académica de altísima calidad y apoyo personal para conseguir los propósitos de esta investigación.
- Karla, Fernanda y Regina, quienes fueron pacientes con los numerosos fines de semana en que dediqué a esta investigación.
- El personal del IIDE, y en especial a quienes colaboran en el DCE, porque me ofrecieron toda la orientación para recorrer la ruta crítica del programa.
- Mis sinodales de tesis, de quienes recibí recomendaciones muy pertinentes sobre cómo mejorar el diseño de esta investigación y comunicar de mejor manera los resultados.
- El Instituto Nacional para la Evaluación de la Educación, en especial a Andrés Sánchez, por haberme dado las facilidades para realizar el estudio y la confianza para analizar en profundidad algunos de sus reactivos.
- El personal de la UAA, en especial a Yolanda, Yahaira, Maru Eugenia y Jesús, por su invaluable apoyo para realizar las entrevistas cognitivas.
- Los docentes y alumnos que aceptaron participar en el estudio cognitivo.
- Mis compañeros de generación, con quienes tuve oportunidad de discutir algunas de las fases de mi investigación, y me ofrecieron recomendaciones muy valiosas.
- El Conacyt, por apoyarme con la beca sin la cual no hubiera podido cursar este programa.



## ÍNDICE DE CONTENIDOS

Índice de tablas .....	xiv
Índice de figuras .....	xvi
RESUMEN .....	xix
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. Antecedentes.....	2
1.2. Planteamiento del problema .....	8
1.3. Objetivos.....	12
1.4. Justificación .....	13
1.5. Los Exámenes para la Calidad y Logro Educativos (Excale) .....	16
1.6. Estructura de la tesis .....	19
<b>2. MARCO TEÓRICO .....</b>	<b>21</b>
2.1. Equidad en las evaluaciones educativas .....	21
2.2. Equidad, sesgo y teoría de la validez en medición .....	26
2.2.1. Estándares para cuidar la justicia en las evaluaciones .....	36
2.2.2. Aspectos de validez diferenciada .....	43
2.3. Funcionamiento Diferencial de los Ítems .....	55
2.3.1. Técnicas no paramétricas para detección del DIF.....	58
2.3.2. Técnicas paramétricas para detección del DIF .....	62
2.3.3. Innovaciones en el análisis DIF .....	65
2.4. Análisis de sesgo en las pruebas .....	67
2.4.1. El diseño universal de las pruebas.....	68
2.4.2. Los protocolos verbales y reportes verbales .....	74
<b>3. MÉTODO .....</b>	<b>78</b>
3.1. Etapa 1. Funcionamiento Diferencial de Ítems .....	79

3.1.1.	Fase 1. Revisión de propiedades psicométricas de los ítems.....	80
3.1.2.	Fase 2. Cálculo del nivel socioeconómico .....	81
3.1.3.	Fase 3. Aplicación del modelo RASCH-DIF por subpoblación según sexo.....	84
3.1.4.	Fase 4. Aplicación del modelo RASCH-DIF por subpoblación según NSE.....	86
3.2.	Etapa 2. Análisis de sesgo mediante reportes y protocolos verbales.....	87
3.2.1.	Fase 1. Diseño.....	88
3.2.2.	Fase 2. Pilotaje .....	97
3.2.3.	Fase 3. Realización de reportes y protocolos verbales.....	98
3.2.4.	Fase 4. Procesamiento de la información.....	103
3.2.5.	Fase 5. Análisis cognitivo de cada ítem .....	106
3.2.6.	Fase 6. Análisis de sesgo mediante el diseño universal de las evaluaciones.....	108
<b>4.</b>	<b>RESULTADOS DEL ANÁLISIS DIF.....</b>	<b>110</b>
4.1.	Propiedades psicométricas bajo la Teoría de Respuesta al Ítem .....	110
4.2.	Funcionamiento Diferencial de Ítems por Sexo .....	114
4.2.1.	DIF Uniforme por Sexo.....	115
4.2.2.	DIF No Uniforme por Sexo.....	125
4.3.	Funcionamiento Diferencial de Ítems por NSE .....	130
4.3.1.	DIF Uniforme por NSE.....	130
4.3.2.	DIF No Uniforme por NSE.....	133
4.4.	Ítems con sospecha de Sesgo.....	140
<b>5.</b>	<b>RESULTADOS DEL ANÁLISIS DE SESGO MEDIANTE PROTOCOLOS VERBALES.....</b>	<b>142</b>
5.1.	Definición imprecisa del constructo .....	144
5.1.1.	Imprecisión en la medición del constructo del ítem SMB-03.....	145

5.1.2. Imprecisión en la medición del constructo del ítem SMC-10 .....	153
5.2. Ítem no accesible o sesgado .....	158
5.2.1. Inaccesibilidad o sesgo del reactivo SMB-01 .....	160
5.2.2. Inaccesibilidad o sesgo del reactivo SMD-04 .....	164
5.3. Procedimientos e instrucciones complejas, poco claras e intuitivas.....	167
5.4. Incomprensibilidad o problemas de legibilidad del ítem.....	170
5.4.1. Incomprensibilidad o problemas de legibilidad del reactivo SMH-11 .....	172
5.4.2. Incomprensibilidad o problemas de legibilidad del ítem SME-09 .....	179
<b>CONCLUSIONES</b> .....	184
<b>REFERENCIAS</b> .....	194
<b>ANEXOS</b> .....	207
Anexo 1 Reactivos por cada forma de Cuadernillo del Excale 09 de Matemáticas. Aplicación 2012 .....	208
Anexo 2. Script para el cálculo del DIF de la subescala SNPA mediante librería TAM de R.....	211
Anexo 3. Formato para Desglosar Procesos Cognitivos a Emplear en la Respuesta a Ítems del Excale-09 de Matemáticas .....	213
Anexo 4. Niveles cognitivos establecidos en el Marco de Referencia del Excale-09 de Matemáticas.....	217
Anexo 5. Guía para la realización de los protocolos y reportes verbales mediante la técnica de pensamiento en voz alta y entrevista concurrente.....	219
Anexo 6. Formato de recolección de datos.....	222
Anexo 7. Carta de Consentimiento .....	224
Anexo 8. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Sentido Numérico y Pensamiento Algebraico</i> del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	225

Anexo 9. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	227
Anexo 10. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	228
Anexo 11. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	229
Anexo 12. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	232
Anexo 13. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	235
Anexo 14. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	238
Anexo 15. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	240
Anexo 16. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	241
Anexo 17. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	242
Anexo 18. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	245

Anexo 19. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	248
Anexo 20. Resumen de reactivos con DIF (uniforme y/o no uniforme) del Excale 09 de Matemáticas en 2012 .....	251
Anexo 21. Análisis de sesgo del reactivo SMB-01 .....	254
Anexo 22. Análisis de sesgo del reactivo SMB-03 .....	263
Anexo 23. Análisis de sesgo del reactivo SMB-09 .....	271
Anexo 24. Análisis de sesgo del reactivo SMB-11 .....	278
Anexo 25. Análisis de sesgo del reactivo SMC-08 .....	284
Anexo 26. Análisis de sesgo del reactivo SMC-10 .....	291
Anexo 27. Análisis de sesgo del reactivo SMD-04 .....	299
Anexo 28. Análisis de sesgo del reactivo SMD-05 .....	306
Anexo 29. Análisis de sesgo del reactivo SMD-11 .....	312
Anexo 30. Análisis de sesgo del reactivo SME-02 .....	319
Anexo 31. Análisis de sesgo del reactivo SME-09 .....	326
Anexo 32. Análisis de sesgo del reactivo SME-11 .....	332
Anexo 33. Análisis de sesgo del reactivo SME-12 .....	339
Anexo 34. Análisis de sesgo del reactivo SMF-02 .....	346
Anexo 35. Análisis de sesgo del reactivo SMG-07 .....	351
Anexo 36. Análisis de sesgo del reactivo SMH-11 .....	357
Anexo 37. Análisis de sesgo del reactivo SMA-05 .....	364
Anexo 38. Análisis de sesgo del reactivo SMG-05 .....	371

## Índice de tablas

Tabla 1. Muestra de alumnos y escuelas de los Excale, según el grado en que se aplican .....	7
Tabla 2. Dominios y subdominios que evalúan los Excale 09 en Matemáticas .....	19
Tabla 3. Etapas de la evolución del concepto de validez .....	34
Tabla 4. Estándares sobre equidad de AERA-APA-NCME, 2014 .....	37
Tabla 5. Estándares sobre equidad del Educational Testing Service, 2014.....	42
Tabla 6. Clasificación de las técnicas de DIF .....	58
Tabla 7. Modelo metodológico del análisis de sesgo .....	79
Tabla 8. Variables empleadas e índices de bondad de ajuste para construir el nivel socioeconómico .....	83
Tabla 9. Cantidad de participantes en el estudio cognitivo, según su nivel socioeconómico y la modalidad de secundaria en la que estudian .....	96
Tabla 10. Valores Infit y Outfit de la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas .....	112
Tabla 11. Valores Infit y Outfit de la subescala <i>Forma, Espacio y Medida</i> del Excale 09 de Matemáticas.....	113
Tabla 12. Valores Infit y Outfit de la subescala <i>Manejo de la Información</i> del Excale 09 de Matemáticas.....	114
Tabla 13. Nivel y sentido del DIF mediante paquetería TAM en la subescala Sentido Numérico y Pensamiento Algebraico, desglosado por Sexo .....	119
Tabla 14. Nivel y sentido del DIF mediante paquetería TAM en la subescala Forma, Espacio y Medida, desglosado por Sexo .....	121
Tabla 15. Nivel y sentido del DIF mediante paquetería TAM en la subescala Manejo de la Información, desglosado por Sexo .....	124
Tabla 16. Resumen de ítems con DIF y NUDIF por Nivel Socioeconómico del Excale 09 de Matemáticas en 2012 .....	141

Tabla 17. Número de reactivos que presentan problemas de sesgo y cantidad de informantes que aportaron evidencias de sesgo, desagregado por tipo.....	143
Tabla 18. Número de entrevistados por ítem que aportaron evidencias sobre una definición imprecisa del constructo, por nivel socioeconómico.....	144
Tabla 19. Número de entrevistados por ítem que aportaron evidencias sobre inaccesibilidad o sesgo, por nivel socioeconómico.....	160
Tabla 20. Número de entrevistados por ítem que aportaron evidencias sobre incomprendibilidad o legibilidad del ítem, por nivel socioeconómico.....	172
Tabla 21. Cantidad de informantes que aportaron evidencias en los ítems diagnosticados con sesgo potencial, desglosado por subescala y tipo de sesgo.....	186
Tabla 22. Cantidad de evidencias de sesgo, por reactivo, tipo de DIF, tipo de sesgo y nivel socioeconómico.....	187

## Índice de figuras

Figura 1. Ejemplo de DIF uniforme (izquierda) y DIF no uniforme (derecha) en las Curvas Características de los Ítems.....	57
Figura 2. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Sentido Numérico y Pensamiento Algebraico</i> del Excale 09 de Matemáticas en 2012, desglosado por Sexo .....	117
Figura 3 Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Forma, Espacio y Medida</i> del Excale 09 de Matemáticas en 2012, desglosado por Sexo.....	120
Figura 4. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Manejo de la Información</i> del Excale 09 de Matemáticas en 2012, desglosado por Sexo.....	122
Figura 5. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Sentido Numérico y Pensamiento Algebraico</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo .....	127
Figura 6 Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Forma, Espacio y Medida</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo .....	128
Figura 7. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Manejo de la Información</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo .....	129
Figura 8 Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Sentido Numérico y Pensamiento Algebraico</i> del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico .....	131



Figura 9 Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Forma, Espacio y Medida</i> del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico.....	132
Figura 10. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala <i>Manejo de la Información</i> del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico.....	133
Figura 11. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Sentido Numérico y Pensamiento Algebraico</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico .....	135
Figura 12. Curvas del Funcionamiento Diferencial de Ítems No Uniforme, según el Nivel Socioeconómico de los alumnos, de ítems SMA_05 y SMG_05.....	136
Figura 13. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Forma, Espacio y Medida</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico .....	138
Figura 14. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala <i>Manejo de la Información</i> del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico .....	139
Figura 15. Reactivo SMB-03 .....	146
Figura 16. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-03.....	148
Figura 17. Operaciones realizadas por E08 para responder ítem SM-03 .....	150
Figura 18. Operaciones realizadas por E05 para responder ítem SM-03 .....	152
Figura 20. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMC-10.....	156
Figura 23. Recorte de un ítem respondido por uno de los alumnos de NSE bajo.....	169

Figura 24. Recorte de un ítem respondido por uno de los alumnos de NSE medio.....	170
Figura 26. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMH-11.....	174
Figura 27. Evidencia del informante E08 para responder el ítem SMH-11.....	176
Figura 28. Evidencia del informante E04 para responder el ítem SMH-11.....	177
Figura 29. Reactivo SME-09 .....	180
Figura 30. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-09.....	181

## RESUMEN

De forma consistente, los resultados de logro educativo para los alumnos de educación básica presentan un patrón: los alumnos de escuelas privadas consiguen los puntajes más altos, enseguida se sitúan los alumnos de escuelas urbanas públicas, después aquéllos que asisten a escuelas de modalidad general asentadas en poblaciones rurales y, al final, los que obtienen los resultados más bajos son los estudiantes de escuelas indígenas y cursos comunitarios, las cuales están ubicadas en las localidades más pequeñas del país y, a su vez, atienden a los alumnos con los niveles socioeconómicos más bajos.

El comportamiento en los resultados señalado el párrafo anterior, se interpretó como una evidencia contundente de que, reconociendo el peso específico que tienen los factores escolares para contribuir al logro académico, todavía una proporción importante de ellos está condicionado por el contexto socioeconómico de los estudiantes y sus familias. No obstante, existe otra razón que podría estar ocasionando que de manera reiterada los alumnos de contextos favorecidos son los que consiguen los puntajes más altos y, por el contrario, los que provienen de contextos socioeconómicos bajos no obtengan resultados satisfactorios: que las pruebas empleadas para evaluar el logro de los alumnos de educación básica en México no midan realmente lo que pretenden medir, es decir que las interpretaciones a partir de los puntajes no sean del todo válidas.

Dada la importancia que en México se cuente con evaluaciones que permitan hacer valoraciones más justas sobre el logro de los estudiantes, se propone como objetivo general de esta investigación aportar, desde un enfoque de equidad de

las pruebas, evidencias de validez de los Excale para tercero de secundaria en Matemáticas, mediante análisis del funcionamiento diferencial de ítems y protocolos verbales.

Los Excale se aplicaron en México desde 2005 hasta 2014 ya que en 2015 fueron sustituidas por las pruebas PLANEA. Estos test fueron diseñados, administrados y calificados por el INEE. Los Excale tuvieron como propósito general “conocer lo que los estudiantes mexicanos aprenden del currículo a lo largo de su educación básica” (Sánchez *et al.*, 2012, p. 7).

Los Excale 09, que se aplicaban a los alumnos de tercero de secundaria, incluían las asignaturas de Español, Matemáticas y Formación Cívica y Ética; y en la versión de 2012 se agregó la asignatura de Ciencias (Martínez-Rizo *et al.*, 2015). En este último año, las pruebas Excale se administraron en total a 94,269 alumnos y 3,590 escuelas secundarias.

El estudio se desarrolló en dos etapas. La primera donde se hicieron los análisis DIF, se desarrolló en cinco fases: 1. Revisión de propiedades psicométricas de los ítems; 2. Cálculo del Nivel Socioeconómico; 3. Análisis DIF por sexo mediante RASCH-DIF; y 4. Análisis DIF por Nivel Socioeconómico. La segunda etapa, donde se realizaron los reportes y protocolos verbales, se llevó a cabo en seis fases: 1. Diseño; 2. Pilotaje; 3. Aplicación de reportes y protocolos verbales; 4. Procesamiento de la información; 5. Análisis cognitivo de cada ítem; y 6. Análisis de sesgo mediante el Diseño Universal de las Evaluaciones. Cada una de las dos etapas del proyecto se enfocó en conseguir distintos tipos de evidencias de validez.

Una primera etapa de esta investigación doctoral consistió en el análisis de cada uno de los cien ítems del Excale de matemáticas para secundaria mediante la técnica RASCH-DIF y la utilización de dos programas estadísticos (Winsteps v. 3.91 y R Statistical v. 3.3.1) que permitió corroborar la concurrencia del diagnóstico. Los resultados de este análisis constataron que los ítems del Excale están libres de sesgo por sexo; en cambio, se identificaron 18 reactivos con DIF severo según el nivel socioeconómico (NSE), de los cuales ocho resultaron más difíciles para los alumnos de NSE alto: SMB\_01, SMF\_02, SMB\_09, SMG\_07, SMB\_11, SMC\_08, SME\_11 y SME\_12; otros ocho ítems fueron más difíciles para estudiantes de NSE bajo: SMB\_03, SME\_02, SMD\_04, SMD\_05, SMH\_11, SMC\_10, SMD\_11 y SME\_09; asimismo, dos reactivos presentaron DIF No uniforme (NUDIF): SMA\_05, SMG\_05.

De la subescala Sentido Numérico y Pensamiento Algebraico (SNPA) se identificaron cuatro ítems con DIF severo (SMB\_01, SMB\_03, SME\_02 y SMF\_02) y dos con NUDIF severo (SMA\_05 y SMG\_05). En la subescala Forma, Espacio y Medida (FEM) se diagnosticaron cinco ítems con DIF severo (SMB\_09, SMD\_04, SMD\_05, SMG\_07 y SMH\_11) y ninguno con NUDIF. De la subescala Manejo de la Información (MI) se encontraron siete ítems con DIF severo (SMB\_11, SMC\_08, SMC\_10, SMD\_11, SME\_09, SME\_11 y SME\_12) y ninguno con NUDIF.

Una vez identificados los ítems con DIF severo, se exploró a través de reportes y protocolos verbales, la presencia de sesgo potencial en esos 18 ítems del Excale 09 de Matemáticas. Para llevar a cabo dicho análisis se utilizaron cuatro de las siete categorías del Diseño Universal de las Evaluaciones (DUE) para analizar la

información recolectada, a saber: 1) definición imprecisa del constructo, 2) ítem no accesible o sesgado, 3) procedimientos e instrucciones complejas, poco claras e intuitivas y 4) incomprendibilidad o problemas de legibilidad del ítem.

Los resultados pueden sintetizarse de la siguiente manera: se identificaron 13 ítems con sesgo en al menos una de las cuatro categorías del DUE a partir de las cuales se les analizó. En la subescala MI es donde se encontraron más ítems (seis) con sesgo potencial, en cambio en la subescala SNPA se diagnosticaron tres reactivos con riesgo de sesgo y en la subescala FEM se identificaron cuatro ítems con una alta posibilidad de sesgo. El tipo de sesgo que más se presentó fue la incomprendibilidad o problemas de legibilidad en ocho ítems, en cambio, ninguno de los reactivos tuvo evidencias de que los procedimientos e instrucciones que se les ofrecieron a los sustentantes para saber cómo registrar sus respuestas fueran complejas, poco claras e intuitivas.

Palabras clave: validez, funcionamiento diferencial de los ítems, protocolos verbales, reportes verbales, entrevistas cognitivas, sesgo, varianza irrelevante al constructo, Excale, equidad en las evaluaciones, validez cultural.

## 1. INTRODUCCIÓN

En las pruebas de logro en México suele aparecer un patrón de resultados en el cual los alumnos que provienen de contextos socioeconómicos limitados tienden a obtener niveles de logro más bajos que aquéllos cuyas familias disponen de mayores recursos socioeconómicos (Backhoff, Bouzas, González-Montesinos, Andrade y Hernández, 2008; Sánchez y Andrade, 2009). Este fenómeno admite al menos dos explicaciones posibles: la primera y más extendida es que las condiciones familiares juegan un papel catalizador en el desempeño escolar de tal manera que al haber más ingresos económicos en una familia se puede dotar de recursos que faciliten un ambiente más propicio para lo que se enseña en la escuela. En cambio, la escasez de ingresos económicos obstaculiza que en esos hogares haya los recursos mínimos que contribuyen al éxito escolar (Backhoff *et al.*, 2008; Blanco, 2007; Treviño, Place, Gempp y Donoso, 2013). La segunda posible explicación es que los ítems que conforman las pruebas con las que se evalúa el rendimiento académico presenten algún tipo de sesgo que obstaculicen que los estudiantes de contextos desfavorecidos puedan demostrar su rendimiento escolar, o bien, que los ítems estén diseñados de tal forma que a los alumnos de contextos económicamente favorecidos les resulte más fácil responderlos correctamente por cuestiones ajenas a lo que la prueba pretende medir.

Relacionada con la segunda posible explicación, en esta investigación doctoral se obtuvieron evidencias de validez del Examen para la Calidad y Logro Educativos (Excale) que se aplicaba en el último grado de la educación secundaria, que permitieran aportar información sobre la eventual presencia de sesgo en sus ítems

ante diferentes grupos de alumnos según el nivel socioeconómico (NSE) en el que se ubican sus familias.

Este capítulo está organizado en cinco apartados. En el primero se presentan a manera de antecedentes las pruebas de logro más empleadas en la educación básica en México. En el segundo apartado se muestra el problema del que surge esta investigación. En el tercero se señala el objetivo general del estudio así como sus objetivos particulares. En el cuarto apartado se muestra como justificación la necesidad de que los Excale dispongan de evidencias de validez con enfoque de equidad y, finalmente, en el quinto se presenta la estructura de la tesis.

### **1.1. Antecedentes**

En el plano internacional, la equidad en las evaluaciones es un tema sobresaliente dentro del campo de la medición educativa desde la década de los sesenta del siglo XX (Camilli, 2006; Cole y Moss, 1989; Gipps y Stobart, 2009), cuando surgieron los movimientos a favor de los derechos civiles y posteriormente en favor de las mujeres. Legisladores, periodistas, educadores, administradores educativos y otros grupos en Estados Unidos, comenzaron a cuestionar si las diferencias en los puntajes obtenidos en las pruebas que se empleaban para seleccionar los estudiantes que ingresaban a la universidad, otorgarles una beca, o bien, colocarlos en centros de educación especial, tenían algún tipo de sesgo que hacía que los jóvenes afroamericanos reiteradamente consiguieran menores puntajes que los blancos (Cole y Moss, 1989).



A partir del gran debate que se generó entre genetistas y ambientalistas, los psicómetras se preocuparon por desarrollar metodologías para el diseño de instrumentos que sean capaces de medir con mayor fidelidad el constructo que se pretende medir. Desde entonces, hubo avances importantes en el campo de la medición con el propósito de generar test más equitativos; uno de los más importantes fue el que Holland y Thayer (1988) nombraron como Funcionamiento Diferencial de los Ítems (DIF, por siglas en inglés), para referirse a ítems que al ser respondidos por sustentantes con el mismo nivel de la habilidad que se deseaba medir, tenían distintas probabilidades de resolverlo de manera correcta.

Después de que Holland y Thayer utilizaran una transformación del procedimiento Mantel-Haenszel para medir el DIF, surgieron muchas técnicas que buscan distinguir entre las diferencias reales de grupos de sustentantes (impacto) y el funcionamiento diferenciado como una potencial muestra de sesgo. Entre las técnicas de DIF más conocidas destacan, además de la propia Mantel-Haenszel y la transformación a escala delta que hicieron Holland y Thayer, el test de Lord que se basa en la Ji cuadrada, la medición de áreas de la diferencia entre las curvas características de los ítems (CCI), el test simultáneo de sesgo en los ítems (SIBTEST), los que se apoyan en la regresión logística, los que utilizan el análisis factorial confirmatorio, los basados en la Teoría de Respuesta al Ítem, los que emplean modelos componenciales, entre otros (Camilli, 1992; Clauser y Mazor, 1998; Fidalgo, 1997; Hidalgo y Gómez-Benito, 2010).

En México, el estudio de la equidad en las pruebas educativas es una línea de investigación en ciernes. Si bien algunas instituciones que realizan evaluaciones de logro mediante pruebas estandarizadas, tales como el Instituto Nacional para la Evaluación de la Educación (INEE) y el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) han llevado a cabo estudios de sesgo, estos no han sido incluidos de manera sistemática en sus reportes técnicos, y sólo a partir de las últimas aplicaciones de las pruebas que aplica el INEE se ha instrumentado como un ejercicio sistemático.

En México la aplicación de pruebas de rendimiento académico a grandes cantidades de alumnos de educación básica es una actividad que se realiza de forma extendida desde la última década del siglo XX. Las primeras pruebas se utilizaron para evaluar el Factor Aprovechamiento Escolar del Programa Nacional de Carrera Magisterial; luego, en 1998 comenzaron a aplicarse las pruebas Estándares Nacionales a alumnos de primaria y secundaria, que constituyen el germen de las evaluaciones estandarizadas de aprendizajes como son en la actualidad (Martínez-Rizo y Blanco, 2010).

En los noventa también comenzó la participación México en evaluaciones internacionales, con el primer estudio del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE, 1996) en el Tercer Estudio Internacional de Matemáticas y Ciencias de la IEA (TIMSS, 1995), y posteriormente con las pruebas del Programa para la Evaluación Internacional de Estudiantes (PISA, por siglas en inglés), bajo la coordinación de la OCDE y cuya primera aplicación se realizó en el año 2000 (Martínez-Rizo y Santos, 2009).

Bajo la responsabilidad del INEE, en 2005 comenzó la aplicación anual de una nueva generación de pruebas nombrada Excale, las que junto con los Exámenes Nacionales del Logro Académico en Centros Escolares (ENLACE), que se iniciaron a aplicar año con año a partir de 2006 bajo la dirección de la SEP, constituyen los dos esfuerzos más importantes para evaluar a gran escala los aprendizajes de los alumnos de educación básica.

Los ENLACE se aplicaban de manera anual y censal a partir de tercer grado de primaria hasta el último grado de secundaria, tanto en el sector público como privado. En primaria cubrían las modalidades general, indígena y cursos comunitarios, mientras que en secundaria lo hacían con las generales, técnicas y telesecundarias. Estaban alineadas al currículo y evaluaban Español, Matemáticas y Formación Cívica y Ética. Otra de sus características es que eran criteriosales, lo que implica que se llega a juicios absolutos sobre el nivel de logro de los alumnos, en lugar de compararlos entre sí; otra implicación es que los resultados de las pruebas criteriosales se interpretan de acuerdo al contenido que se mide y suelen diseñarse para evaluar dominios más específicos, a diferencia de las pruebas referidas a una norma que por lo regular evalúan categorías de competencias (habilidades o conocimientos) más amplias o generales (Backhoff, Peón y Sánchez, 2009).

Por otro lado, los Excale comparten las características de alineación al currículo y que utilizan un criterio para evaluar a los alumnos; sin embargo, presentan otras características diferentes a los ENLACE, tal como lo señalan Martínez-Rizo y Santos (2009) así como Backhoff *et al.* (2009): a) son matriciales, b) se aplican a

una muestra representativa de alumnos, c) emplean aplicaciones rigurosamente controladas a cargo de personal ajeno a la escuela, d) utilizan preguntas cerradas y abiertas, e) incluyen cuestionarios de contexto para obtener información de los mismos alumnos, sus maestros, el director de la escuela y, en su caso, de los padres de familia, sobre numerosas variables del contexto de la escuela misma y del hogar, f) son aplicaciones a un solo grado cada año, conformando un ciclo de cuatro años en los que se aplican pruebas sucesivamente a alumnos de 3° de preescolar, 3° y 6° de primaria y 3° de secundaria, y g) generan reportes de resultados que incluyen análisis complejos de los obtenidos por los alumnos en las pruebas, así como de los factores del hogar y la escuela asociados con ellos. Para el procesamiento de las respuestas de los alumnos a las pruebas matriciales se utiliza la técnica de valores plausibles. Las escalas en que se expresan los resultados, y las de los constructos que se forman a partir de los cuestionarios de contexto, se hacen con base en el modelo de Rasch. Para los análisis se utilizan modelos lineales jerárquicos y de ecuaciones estructurales.

Las muestras de alumnos y escuelas de los Excale son de varias decenas de miles de alumnos y de miles de escuelas, que varían según la conformación del nivel educativo evaluado. Por ejemplo, en las últimas aplicaciones, en el nivel de secundaria se utilizó una muestra más amplia (más de 94 mil estudiantes) y en el nivel de preescolar se evaluaron menos niños (poco más de 14 mil), como puede verse en la Tabla 1.

**Tabla 1. Muestra de alumnos y escuelas de los Excale, según el grado en que se aplican**

Nombre	Grado/Nivel	Último año de aplicación	Muestra alumnos	Muestra escuelas
Excale 00	3° Preescolar	2011	14,256	1,451
Excale 03	3° Primaria	2010	70,434	4,196
Excale 06	6° Primaria	2013	50,602	3,964
Excale 09	3° Secundaria	2012	94,269	3,590

**Fuente:** elaboración propia a partir de informes del INEE

Los Excale incorporaron desde su diseño algunos mecanismos para cuidar que sus ítems no incluyan términos o situaciones que pudieran sesgar los resultados entre hombres y mujeres, o bien, que no contengan palabras que resulten ofensivas o discriminantes para ciertas subpoblaciones (Backhoff, Peón, Sánchez y Andrade, 2006). Sin embargo, los Excale no incorporaron entre sus procedimientos regulares, sino hasta recientemente, estudios de sesgo que involucren a los alumnos a los que está destinada la prueba.

El INEE realizó un par de estudios de sesgo sobre los Excale, uno de ellos empleando DIF y otro con un enfoque más cualitativo, realizado en una población focalizada de estudiantes cuya lengua es el maya (Backhoff, Solano-Flores, Contreras, Vázquez y Sánchez, 2015). Sin embargo, los análisis de sesgo en los que se incluyen los puntajes de los estudiantes o el procedimiento para responder los ítems, no son realizados como una práctica regular y sistemática (Martínez-Rizo *et al.*, 2015), por lo cual se desconoce la magnitud en la que los resultados de dichas pruebas presentan algún tipo de distorsión en grupos de sustentantes como aquellos cuya lengua predominante es distinta al español, provienen de un contexto rural, padecen algún problema visual o auditivo severo, pertenecen a un NSE poco favorecido, entre otros. En síntesis, en las pruebas que miden el logro

en el nivel obligatorio del sistema educativo mexicano, los estudios de sesgo son una práctica poco frecuente, si bien este tipo de análisis resultan indispensables para cuidar la equidad en las evaluaciones educativas; sobre todo en contextos como el mexicano, con una gran diversidad cultural y desigualdad socioeconómica.

Hasta 2014 los Excale se aplicaban al final de cada ciclo educativo de la educación obligatoria. A partir de 2015, la SEP y el INEE sustituyeron los Excale por las pruebas del Plan Nacional para las Evaluaciones de los Aprendizajes (PLANEA), sin embargo, los ítems del Excale aún se siguen utilizando<sup>1</sup>.

## **1.2. Planteamiento del problema**

De forma consistente, los resultados de logro educativo para los alumnos de educación básica presentan un patrón: los alumnos de escuelas privadas consiguen los puntajes más altos, enseguida se sitúan los alumnos de escuelas urbanas públicas, después aquéllos que asisten a escuelas de modalidad general asentadas en poblaciones rurales y, al final, los que obtienen los resultados más bajos son los estudiantes de escuelas indígenas y cursos comunitarios, las cuales están ubicadas en las localidades más pequeñas del país y, a su vez, atienden a los alumnos con los niveles socioeconómicos más bajos (Backhoff, Bouzas, Contreras, Hernández y García, 2007; Backhoff, Bouzas, González-Montesinos,

---

<sup>1</sup> En una comunicación personal, en agosto de 2016, con Andrés Sánchez, Director General de Evaluación de Resultados Educativos del INEE y quien es responsable de las pruebas Planea, señaló que los ítems de Excale que aún no se liberan y que siguen vigentes en las tablas de contenidos de Planea son parte del banco de reactivos del cual se pueden elegir ítems para las pruebas Planea.

Andrade y Hernández, 2008; Díaz y Flores, 2010; Díaz, Flores y Martínez-Rizo, 2006; Flores y Díaz, 2013; Zamudio *et al.*, 2012).

Por ejemplo, en el nivel de secundaria y en la asignatura de Español, mientras que sólo 12% de los estudiantes del sector privado fueron colocados en un nivel por debajo del básico<sup>2</sup>, esa situación ocurrió con 34 y 35% de los alumnos de secundarias técnicas y generales, respectivamente, e inclusive con 50% de los estudiantes de telesecundaria, que estos últimos son los que provienen de familias con contextos socioeconómicos más desfavorecidos. De la misma manera ocurrió con los resultados de Matemáticas donde sólo 25% de los alumnos de secundarias privadas fueron clasificados en un nivel por debajo del básico, mientras que eso sucedió en 51 y 54% de los educandos de escuelas generales y técnicas, respectivamente, y la cifra se incrementó a 62% de los alumnos de telesecundaria (Sánchez y Andrade, 2009). Para el caso de primaria, aunque los porcentajes difieren, permanece el patrón que muestra asociación entre los resultados de logro y el contexto socioeconómico (Backhoff *et al.*, 2008; Zamudio, Díaz Argüero y Lepe García, 2012).

El comportamiento en los resultados señalado los dos párrafos anteriores, se interpretó como una evidencia contundente de que, reconociendo el peso específico que tienen los factores escolares para contribuir al logro académico,

---

<sup>2</sup> Los Excale sitúan a los alumnos en cuatro niveles de logro: avanzado, medio, básico y por debajo del básico. “En términos generales, el nivel por debajo del básico indica carencias importantes en el dominio de los contenidos curriculares, lo cual señala dificultades serias para continuar aprendiendo. El nivel básico representa un dominio elemental para progresar. El nivel medio significa un dominio sustancial de los contenidos, lo cual manifiesta un buen aprovechamiento de los contenidos prescritos en el currículo. Por último, el nivel avanzado implica un dominio óptimo de los contenidos, lo cual evidencia el máximo aprovechamiento respecto de lo previsto” (Sánchez y Andrade, 2009, p. 14).

todavía una proporción importante de ellos está condicionado por el contexto socioeconómico de los estudiantes y sus familias (Backhoff *et al.*, 2008; Blanco, 2007; Treviño *et al.*, 2013). No obstante, existe otra razón que podría estar ocasionando que de manera reiterada los alumnos de contextos favorecidos son los que consiguen los puntajes más altos y, por el contrario, los que provienen de contextos socioeconómicos bajos no obtengan resultados satisfactorios: que las pruebas empleadas para evaluar el logro de los alumnos de educación básica en México no midan realmente lo que pretenden medir, es decir que las interpretaciones a partir de los puntajes no sean del todo válidas.

Como la segunda explicación ante el patrón de resultados de logro remite de forma directa a cuestiones de validez y la equidad en la evaluación, éstos serán conceptos fundamentales a lo largo de este trabajo. Los conceptos se desarrollarán en capítulo siguiente. Por ahora, se puede adelantar de forma sintética que la equidad en la evaluación puede considerarse como la ausencia de sesgo en la medición. Cualquier barrera que impida que un individuo demuestre la habilidad o constructo que se desea medir crearía un sesgo en la interpretación de los puntajes de la prueba y los usos que se hagan de ellos, lo que atentaría contra la justicia en la evaluación.

Por ejemplo, si al construir un test que pretende medir el nivel de habilidad matemática de alumnos de sexto de primaria, se utiliza un formato gráfico en el que el tamaño de letra de los ítems e instrucciones tienen ocho puntos, se utiliza un interlineado sencillo y las hojas del cuadernillo son tamaño media carta, los alumnos con debilidad visual tendrían serios problemas para responder



correctamente a la gran mayoría de los ítems del test, aun cuando tuvieran los conocimientos matemáticos necesarios para hacerlo. A menos que esta prueba quisiera evaluar la agudeza visual (lo que no se plantea), este instrumento contendría elementos de inequidad al volverse innecesariamente más difícil para ciertos grupos poblacionales (los débiles visuales). Así, la injusticia en las evaluaciones ocurre cuando hay barreras innecesarias en grupos de sustentantes para que puedan responder de manera correcta los ítems de una prueba; las barreras innecesarias son aquellos elementos distintos a lo que el examen pretende evaluar y que obstaculizan que un sustentante pueda demostrar a cabalidad sus talentos para enfrentarse a las tareas evaluativas (Basterra, Trumbull y Solano-Flores, 2011).

La equidad de las pruebas no sólo remite al problema de la validez diferenciada para grupos minoritarios, como el ejemplo anterior; también se considera que un test atenta contra la equidad cuando incluye contenido ofensivo o denigrante para ciertos grupos de sustentantes, utiliza palabras o términos que son ajenos a ciertas culturas o familias lingüísticas, los problemas a resolver tienen como contexto una situación que es propia de un grupo poblacional, entre otros (Camilli, 2006; Cole y Moss, 1989)

Si bien hay varios procedimientos para identificar el sesgo en las mediciones educativas, el DIF es una de las herramientas más empleadas con este propósito (McNamara y Roever, 2006). El DIF ocurre cuando dos individuos con el mismo nivel de habilidad tienen diferentes probabilidades de responder correctamente a un reactivo, por el hecho de pertenecer a subgrupos distintos, tales como raza,

estatus socioeconómico, etnicidad, contexto lingüístico, discapacidad, contexto cultural, entre otros (Basterra *et al.*, 2011). Otra de las herramientas que se suele utilizar para identificar el sesgo son las técnicas de pensamiento en voz alta, que se aplican a los sustentantes mediante protocolos verbales (Ericsson y Simon, 1993; Fleck y Weisberg, 2004) para conocer si los conocimientos y habilidades que evocan para resolver los ítems son exclusivamente aquéllos que se pretende medir, o bien, si por el contrario, hay elementos del contexto que inciden para que ciertos grupos de sustentantes respondan erróneamente a un bloque de ítems.

### **1.3. Objetivos**

Dada la importancia que en México se cuente con evaluaciones que permitan hacer valoraciones más justas sobre el logro de los estudiantes, se propone como objetivo general de esta investigación aportar, desde un enfoque de equidad de las pruebas, evidencias de validez de los Excale para tercero de secundaria en Matemáticas, mediante análisis del funcionamiento diferencial de ítems y protocolos verbales. Los objetivos particulares son:

- Documentar los fundamentos teóricos de las evidencias de validez basadas en el aspecto de equidad de pruebas educativas.
- Analizar el grado de funcionamiento diferencial de los ítems de los Excale 09 Matemáticas aplicados en 2012 de acuerdo a dos subpoblaciones de estudiantes creadas a partir de su nivel socioeconómico y su sexo.

- Evaluar el diseño de los ítems del Excale 09 de Matemáticas aplicados en 2012 a través de las categorías de análisis del Modelo del Diseño Universal de las Evaluaciones.
- Identificar sesgo potencial en los ítems del Excale 09 de Matemáticas aplicados en 2012 con un grado alto de funcionamiento diferencial a través de protocolos verbales y técnicas de pensamiento en voz alta.

#### **1.4. Justificación**

El desarrollo de esta investigación tiene al menos tres tipos de aportes: de relevancia social, metodológicos y técnicos. Desde una óptica de relevancia social, este trabajo contribuye a que las evaluaciones de logro que se aplican en educación básica sean más justas, sin importar el NSE de los sustentantes, lo que a su vez permite una mejor toma de decisiones de política pública. Si se considera la creciente importancia que en la literatura internacional (AERA-APA-NCME, 2014; Basterra *et al.*, 2011; ETS, 2002) se le otorga a la validez con un enfoque de equidad (que incluye a la validación cultural y la validación para la atención a la diversidad) y se considera a los Excale como uno de los instrumentos que provee información de logro educativo de la más alta relevancia para la toma de decisiones en el contexto mexicano, resulta indispensable obtener evidencias de validez cultural que permitan conocer el grado en que los usos e interpretaciones hechos a partir de los resultados de estos test coinciden con los propósitos con los que fueron diseñados y, de manera primordial, identificar si existen sesgos culturales en algunos de sus ítems o secciones de esta prueba estandarizada.

Desde un punto de vista metodológico, la realización de este tipo de estudios contribuye a que paulatinamente los Excale, o pruebas derivadas de ellos como las de Planea, se elaboren bajo el *diseño universal* con una atención sistemática a posibles fuentes de sesgo cultural durante todas las etapas de producción de sus ítems y el test en general. Aunque los Excale tienen procedimientos de validación que se incluyen desde el diseño de los reactivos que conformarán las pruebas, el enfoque de equidad fue poco incorporado en dichos procesos (Martínez-Rizo et al., 2015). Es decir, aunque el comité de validación de ítems valora los reactivos de los Excale en lo correspondiente a contenido curricular, problemas técnicos de su construcción y sesgo cultural y de género, éste último aspecto es sobre el que se solicitan menos indicadores de invalidez y se centran sobre todo en cuestiones de género, el uso de vocabulario potencialmente agresivo o denigrante y el uso de regionalismos (Backhoff et al., 2006). De manera adicional a estas fuentes de sesgo, se requieren estudios que contemplen otros grupos poblacionales que por sus características socioeconómicas pudieran estar en desventaja al responder la prueba, por ejemplo, para los estudiantes de niveles socioeconómicos bajos, los que viven en localidades rurales, o bien, aquellos cuya lengua materna es indígena, entre otros.

Desde un punto de vista técnico, reconociendo que el ámbito educativo mexicano la validación con un enfoque de equidad así como los análisis DIF son muy escasos, esta investigación aporta conocimiento científico sobre el grado en que una de las pruebas de logro más extendidas en México, en su versión de Matemáticas para tercero de secundaria, presenta sesgo en sus ítems; es decir, el

presente estudio atiende un hueco en la investigación psicométrica en las pruebas de logro en México al incorporar un procedimiento para analizar las evidencias de validez relacionadas con la equidad y en especial con el uso de Rasch-DIF, el Diseño Universal de las Evaluaciones (DUE) y las técnicas de pensamiento en voz alta en pruebas de alcance nacional.

Si bien se reconoce que el INEE ya ha realizado algunos estudios para identificar DIF, esto no ocurrió de manera sistemática en las evaluaciones de logro que ha desarrollado, sino hasta de manera reciente. Además, las poblaciones que en el INEE se ha explorado DIF son distintas a las que se trabajó en esta investigación. Como se señalan Martínez Rizo y sus colegas:

Aunque el análisis DIF apenas comienza a implementarse en EXCALE (que privilegia el estudio de sesgo de la prueba), éste por lo menos ofrece un reporte con un procedimiento emitido por el software de análisis. EXCALE también cuenta con estudios puntuales de DIF, más cercanos al análisis cualitativo, como es el caso del trabajo realizado sobre una población focal para español y lengua maya (Martínez-Rizo *et al.*, 2015, p. 59)

La ausencia de estudios de validez cultural podría dar lugar a que las decisiones de política pública estén basadas en resultados sesgados que podrían desfavorecer de manera injusta a ciertos grupos sociales. Por ello, en este proyecto se plantea hacer una aportación en pro de una evaluación con enfoque de equidad, específicamente haciendo un análisis de sesgo de los ítems del Excale 09. De manera secundaria, se pretende también que el diseño metodológico aquí empleado sirva como un punto inicial para el resto de

evaluaciones de aprendizaje que realiza el INEE y se contemple la equidad en las evaluaciones de forma sistemática y transversal.

### **1.5. Los Exámenes para la Calidad y Logro Educativos (Excale)**

Aunque no se trata propiamente de un apartado introductorio, en el siguiente presentaré información básica sobre algunas características de los Excale, que permitan contextualizar los análisis que se hacen después.

Los Excale se aplicaron en México desde 2005 hasta 2014 ya que en 2015 fueron sustituidas por las pruebas PLANEA. Estos test fueron diseñados, administrados y calificados por el INEE. Los Excale tuvieron como propósito general “conocer lo que los estudiantes mexicanos aprenden del currículo a lo largo de su educación básica” (Sánchez *et al.*, 2012, p. 7).

Los Excale son pruebas de logro que no pretendieron dar resultados individualizados, sino del sistema en su conjunto, por lo cual se aplicaban de forma muestral, a subpoblaciones representativas de estudiantes de cierto grado académico en cada año. Además, puesto que no se pretendió obtener resultados individuales, los Excale fueron matriciales, es decir, se aplicaban repartiendo las preguntas de la prueba en diferentes cuadernillos de tal manera que cada estudiante sólo responde una parte de la escala completa. Como señalan Sánchez y sus colegas:

los diferentes cuadernillos tienen diferentes grupos de preguntas, de tal manera que cada pregunta está en varios cuadernillos. Se atiende la variedad temática y el

equilibrio en la dificultad de las distintas preguntas de un mismo cuadernillo. Por ende, es posible utilizar las preguntas contestadas por los estudiantes para establecer una sola escala en la cual todos los ítems se relacionan entre sí y conocer cuáles son los contenidos que los alumnos dominan como conjunto y cuáles no (Sánchez *et al.*, 2012, p. 9).

Otra característica de los Excale es que son criteriosales, es decir, tenían un referente fijo desde antes de aplicarse con el que comparan el grado de avance de los estudiantes: el plan de estudios nacional y sus programas. Que las pruebas estén alineadas al currículo nacional implica que sólo incluyen contenidos a evaluar que se encuentran en los programas de estudio. Además, ante la imposibilidad de evaluar todos los contenidos curriculares, los equipos que diseñan el Excale, cuidan que los contenidos curriculares (que incluye además del plan de estudios y sus programas a los libros de texto de los alumnos, los libros para los maestros, los ficheros y materiales que se trabajan en los talleres nacionales, entre otros), estén bien representados en las preguntas utilizadas, tanto en la temática y la extensión como en el enfoque.

Los Excale se aplicaban al final de cada ciclo educativo de la educación obligatoria, y se les denominaba agregando un sufijo que identifique el grado que evalúa. Así, el Excale 00 era para tercero de preescolar, Excale 03 y Excale 06 para tercero y sexto de primaria, respectivamente, así como Excale 09 (para tercero de secundaria) y Excale 12 para evaluar el último grado de la educación media superior. Cada ciclo escolar el INEE aplicaba la prueba a grado escolar

diferente. Así, los Excale 09 fueron aplicados en tres ocasiones: 2005, 2008 y 2012.

Los Excale 09, que se aplicaban a los alumnos de tercero de secundaria, incluían las asignaturas de Español, Matemáticas y Formación Cívica y Ética; y en la versión de 2012 se agregó la asignatura de Ciencias (Martínez-Rizo *et al.*, 2015). En este último año, las pruebas Excale se administraron en total a 94,269 alumnos y 3,590 escuelas secundarias.

El INEE utilizó un método de muestreo de dos etapas para cada uno de los dominios sobre los que hace inferencias. En la primera etapa seleccionaba escuelas y en la segunda a los alumnos. Las escuelas se elegían de forma sistemática a partir de listados ordenados de acuerdo con la cantidad de alumnos que atienden (proporcional al tamaño); los estudiantes de cada escuela se seleccionaban mediante un muestreo aleatorio simple (Juárez, Ramírez y Rodríguez, 2006). Los dominios se conformaban por cada una de las 32 entidades federativas del país así como cuatro de las cinco modalidades bajo las cuales operan las secundarias: generales, técnicas, telesecundarias y secundarias privadas; la secundaria para trabajadores se excluyó de la aplicación de los Excale 09 porque proporcionalmente eran muy pocos alumnos y escuelas y, además, su tendencia en matrícula ha ido disminuyendo.

El Excale 09 de Matemáticas evaluaba tres subdominios: a) sentido numérico y pensamiento algebraico (SNPA), b) forma, espacio y medida (FEM), y c) manejo de la información (MI) y ocho subdominios (ver Tabla 2). La escala completa de



Matemáticas incluía 100 ítems (40 de SNPA, 32 de FEM y 28 de MI). En 2012 se emplearon ocho formas o cuadernillos distintos; cada uno incluyó de 37 a 39 ítems de Matemáticas (ver Anexo 1). Cada ítem tuvo cuatro opciones y los alumnos debían seleccionar sólo una de ellas (Sánchez y Andrade, 2009).

**Tabla 2. Dominios y subdominios que evalúan los Excale 09 en Matemáticas**

Asignatura	Dominio	Subdominio
Matemáticas	Sentido numérico y pensamiento algebraico	Significado y uso de los números
		Significado y uso de las operaciones
		Significado y uso de las literales
Matemáticas	Forma, espacio y medida	Formas geométricas
		Medida
		Transformaciones
Matemáticas	Manejo de la información	Análisis de la información
		Representación de la información

**Fuente:** elaboración propia a partir de las tablas de contenidos del Excale 09

## 1.6. Estructura de la tesis

La tesis está organizada en seis capítulos y dos apartados adicionales. El primero de ellos es donde se ubica este apartado. En el segundo capítulo se desarrolla el marco teórico en el que se sustenta la investigación, el cual está organizado en cuatro subcapítulos. En el primero se presentan el desarrollo actual en torno a la equidad en las pruebas educativas; en el segundo subcapítulo se muestra la relación que hay entre equidad, sesgo y teoría de la validez dentro del campo de la medición; el tercero se dedica a los métodos que se emplean para detectar el DIF en las evaluaciones; y en el cuarto subcapítulo se trata en profundidad el análisis de sesgo en las pruebas.

En el tercer capítulo se presenta la metodología empleada en esta investigación. Ahí se describe, para cada una de las dos etapas del estudio y las fases en las que se desarrolló cada una de las etapas.

En el cuarto capítulo se muestran los resultados empíricos de los análisis de DIF. Este capítulo está organizado en cuatro subcapítulos; en el primero se presenta la información psicométrica de los Excale 09, en el segundo se muestran los análisis DIF por sexo, en el tercero se presentan los análisis DIF por NSE, y en el cuarto los ítems con sospecha de sesgo.

En el quinto capítulo se presentan los resultados del análisis de sesgo. Se organizó en cuatro subcapítulos en cada uno de los cuales se muestran las evidencias más representativas de sesgo, según el tipo de problema que presentaron los ítems

En el sexto capítulo se presentan las conclusiones y se organizó en tres subapartados. En el primero se señalan los logros conseguidos mediante el estudio, en función de cada uno de los objetivos de investigación; en el segundo se muestran las limitaciones de la investigación y en el tercero las recomendaciones para realizar futuros estudios sobre este mismo tema y para los diseñadores de pruebas de logro en gran escala.

Además de los seis capítulos, se incluyen dos apartados adicionales. En el primero se despliegan las referencias utilizadas en la tesis y en el segundo se incluyen los anexos que contienen información pormenorizada para aquellos que deseen conocer a detalle los resultados o replicar el estudio.

## **2. MARCO TEÓRICO**

Este capítulo está organizado en cuatro subcapítulos. En el primero sirve para enmarcar el estudio en el tema de la equidad en las evaluaciones educativas. En el segundo subcapítulo se presenta la relación que existe entre equidad, sesgo y la teoría de la validez en el campo de la medición. En el tercer subcapítulo se desarrollan las técnicas de DIF que más se utilizan, clasificándolas en técnicas paramétricas y no paramétricas. Y en el cuarto se describe cómo dentro de los estudios de análisis de sesgo en las evaluaciones, hay un modelo que sobresale por su flexibilidad y eficacia para interpretar los resultados de este trabajo: el diseño universal de las evaluaciones; además, se describe en qué consiste una de las técnicas que va cobrando cada vez más adeptos para la realización de análisis de sesgo en los que se buscan encontrar evidencias de validez de los procesos de respuesta: las entrevistas cognitivas mediante protocolos verbales.

### **2.1. Equidad en las evaluaciones educativas**

La preocupación por las cuestiones de equidad en las evaluaciones surge en primera instancia en los Estados Unidos en la década de 1960 a partir de los movimientos de derechos civiles y más tarde se refuerza con los movimientos pro-mujer, personas con discapacidad y grupos minoritarios (Camilli, 2006). En las pruebas empleadas en EU para seleccionar a los que ingresarían a las universidades, de forma regular resultaban seleccionados una menor proporción de afroamericanos que presentaban el test (Cole y Moss, 1989; Reynolds y

Suzuki, 2013). Esa situación originó una gran polémica entre ambientalistas y genetistas.

Los genetistas argumentaban que los pobres resultados de los estudiantes (no sólo en educación superior, sino también en educación básica) obedecían a la herencia de una mayor inteligencia de los blancos. Los ambientalistas, por su parte, argumentaban que dichos resultados muy posiblemente se debían a tres factores: 1) el contexto socioeconómico de los grupos afroamericanos, al ser menos favorecido, había también menos estímulos que promovieran un mayor aprovechamiento escolar; 2) la escuela no estaba ofreciendo a los grupos de afroamericanos la calidad suficiente para que ellos consiguieran los conocimientos y habilidades que se requerían para ingresar a la universidad; y 3) que las pruebas que se empleaban para seleccionar a quienes ingresarían a la educación superior era previsible que tuvieran algún tipo de sesgo que favorecía a los grupos caucásicos (Reynolds y Suzuki, 2013).

La utilización como argumento del tercero de los factores (instrumentos con sesgo) puso en el escrutinio el tema de injusticia e igualdad de derechos para grupos minoritarios, pues las pruebas proveían información fundamental para la toma de decisiones que afectaban la vida de individuos, por ejemplo, el acceso a la universidad, la selección y promoción de empleados, los diplomas que se otorgaban en el bachillerato, la ubicación en centros de educación especial, entre otras situaciones. Así, las pruebas dejaron de considerarse como neutrales y objetivas, sino que podían estar sesgadas hacia alguno de los grupos poblaciones,

que usualmente se asumía que favorecían a los hombres, blancos y de clase media (Gipps y Stobart, 2009).

A partir de ello, hubo un gran desarrollo en torno al concepto de la equidad o injusticia en las evaluaciones, que fue tomando por un lado un cariz ético y legal y, por otro, uno propio del campo de la medición. El involucramiento de cortes, jueces, periodistas, tomadores de decisiones en política educativa, así como los propios psicómetras, generó números debates en torno a la equidad de las evaluaciones, lo que luego desencadenó que ciertos términos tuvieran múltiples acepciones, que en ocasiones resultaban muy disímiles (Camilli, 2006).

El término sesgo se solía equiparar al concepto de injusticia o inequidad. Y se usaba como sinónimo de injusto, malo, indeseable o desigual (Fidalgo, 1997). Buscando un término más neutro, que remitiera sólo a cuestiones de medición, en 1982 Angoff propone utilizar el término *discrepancia de los ítems*. Fue hasta 1988 cuando Holland y Tyler propusieron el término *funcionamiento diferencial de los ítems* para desplazar al sesgo en el análisis de reactivos. Ese término es el que en la actualidad todavía se utiliza para aludir a cuestiones estadísticas de los ítems. Así, a partir de entonces, la utilización de sesgo se emplea en un sentido general que evoca a asuntos éticos y legales, mientras que Funcionamiento Diferencial de Ítems (DIF, por sus siglas en inglés) se reservó para el campo de la medición, y cuando en esta disciplina se utiliza el concepto de sesgo se concibe como un:

... error sistemático en la estimación de un valor. Un test sesgado es aquel que sistemáticamente sobreestima o subestima el valor de la variable que intenta medir.

Si el sesgo ocurre como una función de una variable nominal cultural, como la etnicidad o el género, se dice que el test presenta sesgo cultural. (Reynolds y Suzuki, 2013, p. 83).

Para detallar más esta cuestión, vale la pena recordar una precisión que hace Martínez-Rizo (2016): la evaluación consiste en la emisión de un juicio en el que se compara la medición de un constructo o una variable con referente establecido. La injusticia puede ocurrir al llevar a cabo la medición (cuestión psicométrica), pero también cuando se define el estándar (cuestión de la prudencia), y además, cuando se emite el juicio (cuestión ética). Si en cualquiera de estos tres elementos no se consideran los usos que se darán a los resultados de una evaluación, para cada uno de los grupos de población a los que está destinada, es muy probable que haya injusticias o sesgos.

El análisis de la equidad en un principio se circunscribía a cuestiones raciales y de género, posteriormente se amplió a otras situaciones como: grupos con alguna discapacidad, diversidad lingüística, grupos minoritarios, acreditación y colocación, entre otros. La psicometría propuso mecanismos para evitar contenido en las pruebas o sus ítems que pudiera resultar ofensivo para esos grupos (u otros en los que se utilizarán los resultados de un test), minimizar la insensibilidad cultural, la creación de acomodaciones para la administración de las pruebas, así como incrementar la accesibilidad de los sustentantes.

Un concepto estrechamente relacionado con el de equidad en la evaluación es el de validez cultural. “La evaluación, después de todo, es una actividad social que sólo puede ser entendida dentro del contexto cultural, económico y político en el

que opera” (Gipps y Stobart, 2010, p. 56). En relación a las evaluaciones de aprendizajes, la cultura influye en la forma en que los alumnos interpretan y responden a las actividades planteadas en las pruebas (Solano-Flores, 2011; Solano-Flores y Nelson, 2001).

La falta de consideración de las características socioculturales de las poblaciones a las que están dirigidas las pruebas puede llevar a conclusiones imprecisas e inválidas sobre su desempeño. Lo anterior hace evidente la necesidad de incorporar en todo el proceso evaluativo “las influencias culturales que moldean la manera en la cual los estudiantes interpretan y resuelven problemas...” (Solano-Flores y Nelson, 2001, p. 554).

Solano-Flores y Nelson (2011) definen la validez cultural como:

...la eficacia con la que [...] la evaluación aborda la influencia socio-cultural que da forma al pensamiento del estudiante y las maneras en que los estudiantes les dan sentido a [...] los reactivos y sus respuestas. Las influencias socio-culturales incluyen el conjunto de valores, creencias, experiencias, patrones de comunicación, estilos de enseñanza y aprendizaje, epistemologías inherentes al contexto cultural de los estudiantes y las condiciones socioeconómicas que prevalecen en sus grupos culturales (citados en Basterra *et al.*, 2011, p. 3).

La validez cultural implica considerar la diversidad de contextos de quienes responderán los instrumentos de evaluación como una posible fuente de varianza irrelevante al constructo que se desea medir, de tal manera que los puntajes obtenidos en una prueba no se vean sesgados por aspectos ajenos al interés de la evaluación (como el contexto social, cultural, económico, etc.). Como puede verse,

los asuntos de equidad están muy relacionados con la validez en la medición. En el siguiente apartado se desarrolla la manera en que interactúan estos dos conceptos en la medición.

## **2.2. Equidad, sesgo y teoría de la validez en medición**

La noción de validez es una de las más fundamentales de la psicometría y, en general, de la investigación en el campo de las ciencias sociales y de la conducta, pero, a la vez, una de las más complejas. La noción surgió hacia 1920, en relación con los primeros esfuerzos por extender la metodología de las pruebas de inteligencia a la evaluación de los aprendizajes y a la medición de actitudes. Para mediados del siglo XX, en el marco de la teoría clásica de la medición, la noción de validez, con la de confiabilidad, había alcanzado ya su lugar como uno de los conceptos centrales del campo.

En las décadas siguientes la noción se fue precisando, primero con la distinción entre validez de criterio (concurrente o predictiva, convergente o discriminante), validez de contenido y validez de constructo, y luego con la concepción unificada propuesta por Messick. La evolución del concepto se aprecia en los capítulos sobre validez en las ediciones de la principal obra de referencia del campo, *Educational Measurement* (Cronbach, 1971; Cureton, 1951; Kane, 2006; Messick, 1989) y en las versiones de los estándares sobre validez de las principales agrupaciones profesionales (Cfr. AERA-APA-NCME 1974, 1985, 1999a y 2014).

Newton y Shaw (2014) identifican cinco etapas en la evolución de la teoría de la validez. Aquí se recuperan las últimas cuatro. Así, durante la etapa conocida como



*Cristalización* (de 1921 a 1951), el concepto de validez se enfocó en la predicción de un criterio específico, que posteriormente se conocería como validez predictiva, aunque también durante esta época se realizaron estudios de validez concurrente; ambos acercamientos se realizaron utilizando métodos correlacionales, salvo que los primeros usaron dos momentos distintos para la recolección de la información, mientras que en los estudios de validación concurrente se recogía simultáneamente datos tanto de la medida del test como del criterio con el que se comparaba. Algunas definiciones representativas de la primera etapa son: en 1937 Garrett señalaba que “la validez de un test es la fidelidad con la que mide lo que pretende medir” y Bingham definía la validez como “la correlación de las puntuaciones de un test con alguna otra medida objetiva de lo que el test quiere medir” (citados en Lissitz, 2009, p. 22). Por su parte, Guilford en 1946 la concebía así: “en un sentido general, una prueba es válida para cualquier cosa con la que se correlaciona” (citado en Messick, 1989, p. 18).

A pesar de esta visión tan empirista, desde esos tiempos ya se reconocían elementos del proceso de validación que siguen vigentes hasta hoy, por ejemplo, el mismo Cureton (1951) señalaba que la pregunta fundamental de la validez de las pruebas es “qué tan bien ejecuta la función para la que fue empleada” (1951, p. 621). Lo que implica que la misma prueba puede ser usada con diferentes propósitos, y su validez puede ser alta para alguno, moderada para otro y baja para un tercer propósito. Por tanto, no se puede etiquetar a una prueba en general como de alta, moderada o baja validez, sino sólo en el contexto de ciertos fines.

En esta época ya era nítida la distinción entre validez y confiabilidad, aunque se reconocía que estaban relacionadas; Cureton (1951) señalaba que la segunda se refiere a la exactitud y consistencia con la que se mide cualquier cosa tomando en cuenta el grupo con el que se utilizará la información. En cambio, para que una prueba sea válida o que sirva adecuadamente a sus propósitos debe medir “algo” con un razonable grado de confiabilidad, y además, ese “algo” debe ser muy similar a las operaciones que se están usando para medirlo. Lo que Cureton advertía es que la validez de un instrumento si bien depende de su relevancia y su confiabilidad, la eficacia de la prueba no es una función de la confiabilidad de las puntuaciones de criterio, sino del grado de consistencia entre los procesos que se desean medir y los que tienen que realizar los sujetos cuando se enfrentan a los instrumentos de medición.

Al inicio de la etapa de la *fragmentación* (1952-1974) aparece lo que en ese tiempo se conoció como validez de contenido, en respuesta a lo restrictivo que les pareció a algunos críticos el hecho de que la validez únicamente se considerara como una correlación con conductas observables consideradas como criterio, que resultaba insuficiente para pruebas donde sus propios puntajes son el criterio a considerar. Ejemplo de ello son los test de rendimiento académico o trastornos de la conducta. Así, la validez incorporó la faceta *de contenido*, donde a partir de expertos se evaluaba si el test representaba y cubría de forma suficiente el constructo que pretendía medir.

En 1954 la *American Psychological Association* (APA) publicó recomendaciones técnicas para la psicología y entre ellas incluyó 19 estándares para preservar la

validez. Por su parte, en 1955 la *American Educational Research Association* (AERA) generó sus propios estándares para el área de la educación. En los inicios de la segunda época se distinguían cuatro tipos de validez: de contenido, predictiva, concurrente y de constructo. En 1966 la APA y la AERA publicaron juntas un solo texto con una versión revisada y aumentada de los criterios anteriores, titulado *Estándares para la Evaluación Educativa y Psicológica*. Ahí se conforma lo que hasta hoy todavía algunos psicómetras denominan la trinidad de la validez, refiriéndose a lo que en ese entonces se le conoció como tres tipos de validez: de contenido, de constructo y de criterio (ésta última podría ser predictiva o concurrente).

La *validez de criterio* es la que se consigue al comparar las puntuaciones de una prueba o predicciones hechas a partir de ellas, con una variable externa o criterio que provee una medición directa de las características de la conducta en cuestión. La *validez de contenido* evalúa qué tan bien el contenido de una prueba muestrea los tipos de situaciones o materias acerca de las cuales se harán las inferencias. Finalmente, la *validez de constructo* es evaluada al indagar las cualidades psicológicas o variables que realmente mide un instrumento de medición. En el concepto de validez que propusieron la APA y la AERA en 1966 prevalece la idea de que la validación debía considerar aspectos descriptivos y teóricos con requerimientos procedimentales y lógicos que respaldasen los argumentos de validación. Es decir, la validación de un instrumento implica la integración de distintos tipos de evidencia. Las pruebas son hechas con múltiples propósitos y es muy raro que un criterio sea el verdaderamente principal. El autor del capítulo

dedicado a la validación de la segunda edición del *Educational Measurement*, señalaba que la mayoría de las mediciones serían inviables si uno necesitara diferentes pruebas para cada decisión o propósito (Cronbach, 1971).

No obstante que los *Estándares* de 1996 y 1974 promovieron una visión trinitaria de la validez, muchos autores y usuarios interpretaron erróneamente que los tests podían validarse a través de cualquiera de los procedimientos y que era suficiente utilizar sólo uno de ellos, lo que derivó en que durante esta época se inventaran una enorme variedad de adjetivos o variantes de la validez, distintos a los propuestos por los *Estándares*. En una presentación de Newton (2013), recupera de esa época casi un centenar de “tipos” de validez, donde aparecen por ejemplo, validez factorial, validez *in-situ*, validez inferencial, validez del proceso de calificación, validez funcional, validez interpretativa, etc.

En síntesis, la etapa de la fragmentación se sigue caracterizando, salvo las aportaciones de Cronbach, por su empirismo y escasa teorización; sin embargo, uno de los avances más importantes que se lograron durante esa época fue considerar que la validez no es una propiedad del test, sino que será en función de lo adecuadas que resulten las inferencias hechas a partir de los puntajes de las pruebas o instrumentos de medición; asimismo, se logró distinguir que la validación es un proceso y que la validez es una propiedad de las inferencias, la cual admite grados, en función de la cantidad de evidencias que las soporten.

En la etapa de la *reunificación* (1975-1999) hubo avances muy importantes en la conceptualización de la validez. Así, la validez de constructo se consideró el

concepto unificador sobre el cual se adquieren diferentes tipos de evidencias de validez: contenido, procesos de respuesta, estructura interna, relación con otras variables y, usos y consecuencias. Desde una visión general, la validez se concibe como un resumen inductivo de las evidencias existentes tanto de las consecuencias potenciales a partir de la interpretación de las puntuaciones de una prueba como de sus usos. Por lo tanto, lo que debe ser validado no es la prueba o el dispositivo que se utiliza para la observación, sino las inferencias derivadas de las puntuaciones de las pruebas u otros indicadores, como las inferencias o interpretaciones del significado de una puntuación y las implicaciones que conlleva dicha interpretación.

Desde esta etapa la validez es considerada como una cuestión de grado, no de todo o nada. Además, con el tiempo, la evidencia de validez existente se fortalece o debilita por nuevos hallazgos y, además, las proyecciones de las posibles consecuencias sociales de las evaluaciones se transforman a partir de la evidencia sobre las consecuencias reales en la actualidad y las condiciones sociales cambiantes. Entonces, inevitablemente, la validez es una propiedad en evolución y la validación es un proceso continuo. Dado que la evidencia siempre es incompleta, la validación es fundamentalmente una cuestión de tomar en cuenta los usos actuales de una prueba así como la investigación más reciente para mejorar el entendimiento de lo que significan las puntuaciones de un instrumento de medición.

Validar una inferencia interpretativa es comprobar el grado en que múltiples tipos de evidencia son consonantes con la inferencia, mientras que inferencias

alternativas están menos soportadas. Validar una inferencia de acción requiere validar no sólo el significado de cierto puntaje en un instrumento de medición sino el valor de las implicaciones y de los resultados de las acciones, especialmente evaluando la relevancia y utilidad de la puntuación en una prueba para un propósito específico, así como las consecuencias sociales de usar una puntuación para la toma de decisiones (Messick, 1989).

En la última etapa (de 2000 a la fecha), conocida por la *deconstrucción* de la validez pueden destacarse varios aspectos en la evolución del concepto de validez. Uno es el cambio en el énfasis, de numerosas formas específicas validez de criterio, a un pequeño número de tipos de validez, luego a una concepción unitaria de la noción y, finalmente, a distinguir cinco tipos de evidencias de validez necesarias. Otro es el cambio en el enfoque central de validación, primero centrado en la predicción y luego en la explicación, en el sentido de que la utilidad, relevancia e importancia de la predicción no pueden ser evaluadas en ausencia de una interpretación fundamentada en datos empíricos consistentes.

Para vincular el proceso de validación al de un programa de investigación científica, Kane (2001) propuso un enfoque basado en argumentos que el diseñador genera en función de los propósitos que persigue con la prueba, plantea hipótesis de trabajo y luego obtiene evidencias que respalden o refuten los argumentos e hipótesis que se planteó.

Aunque las ideas prevalecientes en la actualidad entre los estudiosos del campo y los diseñadores de pruebas han sido las establecidas en la tercera etapa, desde

inicios del siglo XXI, surgió un grupo de autores como Borsboom, Mellenbergh y van Haerden (2004), Lissitz y Samuelsen (2007) que aunque comparten varios de los postulados básicos de la visión más extendida de validez, son críticos sobre los alcances que tiene la concepción actual de la validación y difieren en aspectos epistemológicos y ontológicos.

Para estos autores, la validez sí es una propiedad del test y la definen a partir de sus características internas (confiabilidad y contenido). Aunque reconocen la importancia de otras fuentes de evidencia, para ellos las más importantes son las relacionadas con la validación de contenido, los procesos de respuesta y la confiabilidad, señalando que las otras fuentes son ajenas al proceso de validación y escapan del campo de acción de los diseñadores de los test.

Así, mientras que para los autores de la visión predominante (Kane, 2013; Messick, 1989) opinan que el diseñador de las pruebas debería cuidar que los usos que se hagan con los resultados de los test sean acordes con los propósitos para los que fue diseñado, Borsboom, Mellenbergh y van Haerden (2004), Lissitz y Samuelsen (2007) rechazan de manera tajante dicha postura, argumentando que esas acciones escapan del campo de responsabilidades de los diseñadores de instrumentos de medición y, por tanto, no afectan la validez de un instrumento.

En la Tabla 3 se presenta de forma resumida cómo evolucionó el concepto de validez, los principales autores de cada etapa y las principales aportaciones que surgieron en cada una a la teoría de la validación.

**Tabla 3. Etapas de la evolución del concepto de validez**

Etapa	Concepción de Validez	Principales autores	Aportes principales a la teoría de la validación
1. Cristalización (de 1921 a 1951)	La validez centrada en la validez predictiva. La correlación entre las puntuaciones observadas del test con las puntuaciones 'verdaderas' del criterio.	Garrett Guilford Cureton	<ul style="list-style-type: none"> <li>- La validez se distinguió de la confiabilidad.</li> <li>- Una prueba puede tener varios propósitos y su validez puede ser alta para unos de ellos y baja otros de esos propósitos.</li> </ul>
2. Fragmentación de la validez (1952-1974)	Desde los teóricos, la validez podría ser de tres tipos: de contenido, de constructo y de criterio (ésta última podría ser predictiva o concurrente); sin embargo, entre los diseñadores de instrumentos y autores inventaron, de forma errónea, una enorme variedad de adjetivos de la validez. Newton recuperó casi un centenar "tipos" de validez usados en esta época.	Cronbach Meehl	<ul style="list-style-type: none"> <li>- Surge la validez de contenido y prevalece como una faceta importante de la validez, aunque esa época se concebía como un tipo y no como una dimensión.</li> <li>- Surge la validez de constructo, aunque todavía como un tipo más de validez y no con la importancia que cobraría con posterioridad.</li> <li>- La validación de un instrumento implica la integración de distintos tipos de evidencia. Las pruebas son hechas con múltiples propósitos y es muy raro que un criterio sea el verdaderamente principal.</li> </ul>
3. Reunificación (1975-1999)	La validez es un concepto unitario que siempre refiere al grado en que la evidencia empírica y el fundamento teórico apoyan lo adecuado de las interpretaciones y acciones realizadas a partir de las puntuaciones de un instrumento	Messick Embretson Cronbach	<ul style="list-style-type: none"> <li>- Ante la gran fragmentación de la validez, los principales teóricos argumentan que toda validez es de constructo y que el resto de dimensiones forman parte de esta primera dimensión.</li> <li>- La validación como un proceso de investigación científica, que tendría que ocurrir en paralelo al diseño de instrumentos y posterior a ello.</li> <li>- La validez no como una propiedad de los test, sino de las interpretaciones de los resultados y de sus usos.</li> </ul>
4. Deconstrucción (2000 a la fecha)	El grado en el que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba y los usos que se pretende hacer de ellos. Se tendría que obtener evidencias de cinco fuentes: de contenido, de procesos	<p>Visión predominante: Kane Messick</p> <p>Visión crítica: Borsboom Lissitz</p>	<ul style="list-style-type: none"> <li>- Las consecuencias y usos como dimensión de primera importancia.</li> <li>- La equidad en las evaluaciones como un aspecto central del proceso de diseño de los instrumentos.</li> <li>- Metodología de validación con un Enfoque de Validación Basado en Argumentos.</li> </ul>



Etapa	Concepción de Validez	Principales autores	Aportes principales a la teoría de la validación
	de respuesta, estructura interna, relaciones de los puntajes del instrumento con otras variables y consecuencias de la evaluación		

**Fuente:** García-Medina, Martínez-Rizo, Cordero y Caso (2016). *Evolución del concepto de validez en la medición educativa*, a partir de Newton y Shaw (2014).

Uno de los elementos del proceso de validación que inició desde la etapa de la reunificación y que adquirió cada vez más fuerza durante la etapa de la deconstrucción, es la justicia y equidad en las evaluaciones, o *fairness*. La equidad en las evaluaciones educativas y psicológicas<sup>3</sup> se entiende como:

la atención a las características individuales y contextos de las pruebas de tal manera que los puntajes lleven a interpretaciones válidas para los destinatarios...una prueba es equitativa, cuando refleja el mismo constructo para todos los sustentantes, y los puntajes tienen el mismo significado para todos los individuos de la población objetivo (AERA-APA-NCME, 2014, p. 50).

La equidad en la evaluación es un aspecto central del proceso de validación que debería cuidarse en todas las etapas de desarrollo de una prueba, la interpretación de los resultados y los usos que se hagan de ellos, es decir, es un elemento que debería ser transversal (ETS, 2002, 2009, 2014). Así, la idea central de la equidad en la evaluación consiste en identificar y eliminar la varianza irrelevante de constructo para maximizar el desempeño de cualquier examinado,

<sup>3</sup> En la literatura internacional se le llama *fairness*, término que es traducido como justicia o equidad; en este documento se privilegiará el uso del segundo de estos últimos términos.

de tal manera que los puntajes sean equiparables entre los distintos sustentantes de la prueba.

Para promover y valorar la justicia en las evaluaciones se definieron estándares que guían los diferentes procesos y acomodaciones que se incluyen en las evaluaciones estandarizadas, los cuales se presentan en el siguiente apartado.

### **2.2.1. Estándares para cuidar la justicia en las evaluaciones**

En el desarrollo de las pruebas, se han establecidos estándares que pretenden garantizar la equidad en las evaluaciones. Dos de las normativas más difundidas sobre ese tenor, son las publicadas por AERA-APA-NCME, y por el Educational Testing Service (ETS).

#### *Estándares de AERA-APA-NCME para cuidar la equidad en las evaluaciones*

En la versión más reciente de los Estándares (AERA *et al.*, 2014), los relacionados con la justicia de las evaluaciones se dividen en cuatro grupos (ver Tabla 4). El primero se refiere al diseño de la prueba, desarrollo, administración y procedimientos de calificación que minimicen las barreras para una interpretación válida del mayor rango posible de individuos y subgrupos relevantes. En la definición puede verse que los estándares que se encuentran en este grupo, tienen que ver con las diferentes fases por las que atraviesa una evaluación. Comprende, por ejemplo, la definición clara del constructo a medir y las características de los individuos y subgrupos a quienes se pretende aplicar las evaluaciones, la definición de consignas e ítems en los que los estímulos no necesarios sean erradicados de tal manera que se maximice el acceso a los

sustentantes para manifestar su posición con respecto al constructo medido, la realización de estudios piloto en los que se incluyan subgrupos relevantes, y la provisión de un tratamiento comparable durante la administración y calificación de la prueba.

**Tabla 4. Estándares sobre equidad de AERA-APA-NCME, 2014**

Grupo	Estándar
<p>Grupo 1. Procedimientos de diseño, desarrollo, administración y puntuación de pruebas que minimizan las barreras para hacer interpretaciones válidas de los puntajes del rango más amplio posible de individuos y subgrupos relevantes</p>	<p>Estándar 3.1. Los responsables del desarrollo, revisión y administración de la prueba deberían diseñar todos los pasos del proceso de evaluación para promover interpretaciones válidas de los puntajes con respecto al uso que se pretendía hacer de ellos, en el más amplio rango posible de individuos y subgrupos relevantes en la población objetivo de evaluación.</p>
	<p>Estándar 3.2. Los diseñadores de la prueba son los responsables de desarrollar un test que mida el constructo que se pretende medir y de minimizar que los sustentantes sean afectados por características irrelevantes al constructo, tales como las lingüísticas, comunicativas, cognitivas, culturales, físicas u otras características.</p>
	<p>Estándar 3.3. Los responsables del desarrollo de una prueba deberían incluir subgrupos relevantes en la validación, la confiabilidad/precisión y otros estudios preliminares que se usan en la construcción de un test.</p>
	<p>Estándar 3.4. Los sustentantes deberían recibir un tratamiento comparable durante la administración de la prueba y el proceso de calificación.</p>
	<p>Estándar 3.5. Los diseñadores de la prueba deberían especificar y documentar las provisiones que han hecho para probar los procedimientos de administración y calificación para eliminar las barreras irrelevantes al constructo de todos los subgrupos relevantes de la población que toma el test.</p>
	<p>Estándar 3.6. Cuando haya evidencia creíble que indique que los puntajes de los exámenes pueden diferir en significado para los subgrupos relevantes en la población que se pretende estudiar, los desarrolladores de pruebas y/o usuarios son responsables de examinar la evidencia para validar las interpretaciones de los puntajes en función de los usos previstos para los individuos de esos subgrupos.</p>
<p>Grupo 2. Validez de la puntuación de la prueba. Interpretaciones para usos previstos para la población examinada planeada.</p>	<p>Lo que constituye una diferencia significativa en las puntuaciones de los subgrupos y qué acciones se realizan en respuesta a tales diferencias puede ser definido por las leyes aplicables.</p>
	<p>Estándar 3.7. Cuando la evidencia de validez basada en un criterio se utiliza como fundamento para hacer predicciones sobre un desempeño futuro a partir de la puntuación de una prueba, y el tamaño de la muestra es suficiente, los desarrolladores del test y/o los usuarios son responsables de evaluar la posibilidad de una predicción diferencial para subgrupos relevantes con base en evidencia previa o teoría que sugiere esa predicción diferenciada.</p>
<p>Grupo 3. Acomodaciones para eliminar</p>	<p>Estándar 3.8. Cuando las pruebas requieren la calificación de respuestas construidas, los diseñadores y usuarios deberían recoger y reportar evidencias de validez sobre la interpretación de los puntajes de acuerdo a los usos previstos, para subgrupos relevantes de las poblaciones objetivo de examinados.</p>
	<p>Estándar 3.9. Los desarrolladores y/o usuarios de las pruebas son responsables de desarrollar y proveer acomodaciones a los test, cuando sea apropiado y factible, para eliminar barreras irrelevantes al constructo que, de</p>

Grupo	Estándar
<p>barreras irrelevantes al constructo y apoyar interpretaciones válidas de los puntajes para los usos previstos.</p>	<p>otra manera, podrían interferir con la habilidad de los examinandos de demostrar su dominio sobre los constructos objetivo.</p> <hr/> <p>Estándar 3.10. Cuando en la prueba se permiten acomodaciones, los desarrolladores y/o usuarios del test son responsables de documentar las disposiciones estándar para usar las acomodaciones y para monitorear su implementación apropiada .</p> <hr/> <p>Estándar 3.11. Cuando se cambia una prueba para eliminar las barreras de accesibilidad al constructo que se está midiendo, los desarrolladores y/o usuarios del test son responsables de obtener y documentar evidencias de validez de las interpretaciones de los puntajes para los usos previstos de la prueba modificada, cuando el tamaño de la muestra lo permita.</p> <hr/> <p>Estándar 3.12. Cuando una prueba es traducida y adaptada de un idioma a otro, los desarrolladores y/o usuarios del test son responsables de describir los métodos utilizados para establecer la adecuación de la adaptación y documentar la evidencia empírica o lógica para la validez de las interpretaciones de los puntajes en función del uso que se les pretende dar.</p> <hr/> <p>Estándar 3.13. Una prueba debe ser administrada en el idioma más relevante y apropiado para el propósito de la prueba.</p> <hr/> <p>Estándar 3.14                      Cuando la evaluación requiere el uso de un intérprete, el intérprete debe seguir procedimientos estandarizados y, en la medida de lo posible, ser lo suficientemente fluido en el idioma y contenido de la prueba y en el idioma nativo y cultura del examinado, para traducir la prueba y los materiales de relacionados a ella y que explique las respuestas al test del examinado, según sea necesario.</p>
<p>Grupo 4.                      Dispositivos de seguridad contra interpretaciones inadecuadas de los puntajes en función de los usos previstos</p>	<p>Estándar 3.15. Los desarrolladores de pruebas y editores que afirman que se puede dar un descanso a los examinandos de subgrupos específicos son responsables de proporcionar la información necesaria para soportar que las interpretaciones de los puntajes de la prueba son apropiadas para sus usos previstos con los individuos de estos subgrupos.</p> <hr/> <p>Estándar 3.16. Cuando investigaciones creíbles indiquen que las puntuaciones del test de algunos subgrupos relevantes se ven afectadas de manera diferenciada por características irrelevantes al constructo de la prueba o de los examinados, cuando sea legalmente admisible, los usuarios de las pruebas deberían utilizar un descanso sólo para esos subgrupos para los que hay suficientes evidencias de validez de que apoyen que las interpretaciones de los puntajes son para los usos previstos.</p> <hr/> <p>Estándar 3.17. Cuando los puntajes agregados son públicamente reportados para subgrupos relevantes -por ejemplo, hombres y mujeres, individuos de diferente estatus socioeconómico, individuos que difieren por raza / etnia, individuos con diferentes orientaciones sexuales, individuos con diversos contextos lingüísticos y culturales, personas con discapacidades, niños pequeños o adultos mayores-, los usuarios de las pruebas son responsables de proporcionar evidencia de comparabilidad y de incluir advertencias, siempre que investigaciones o teorías creíbles indiquen que los puntajes de las pruebas pueden no tener un significado comparable a través de estos subgrupos.</p> <hr/> <p>Estándar 3.18. Al evaluar individuos con propósitos de diagnóstico y/o colocación en un programa especial, los usuarios de las pruebas no deberían usar los puntajes de las pruebas como los únicos indicadores para caracterizar el funcionamiento, la competencia, las actitudes y/o las predisposiciones de un individuo. En su lugar, deberían utilizarse múltiples fuentes de información, considerar explicaciones alternativas sobre el desempeño en la prueba y el juicio profesional de alguien familiarizado con la prueba debería ser incluido en la decisión.</p>

Grupo	Estándar
	<p>Estándar 3.19. En casos en los que la misma autoridad es responsable tanto de la provisión de planes de estudio como de decisiones de alto impacto basadas en la evaluación del dominio curricular de los sustentantes, los examinados no deberían sufrir consecuencias negativas permanentes si la evidencia indica que ellos no han tenido la oportunidad de aprender el contenido de la prueba.</p>
	<p>Estándar 3.20            Cuando un constructo puede medirse de diferentes maneras que sean iguales en su grado de representación del constructo y validez (incluida la ausencia de varianza irrelevante al constructo), los usuarios de la prueba al decidir qué test y/o punto de corte utilizar deberían considerar, entre otros factores, evidencia de las diferencias de los subgrupos en las puntuaciones promedio o en el porcentajes de examinados cuyas puntuaciones exceden los puntos de corte.</p>
<p><b>Fuente:</b> traducción propia a partir de AERA-APA-NCME (2014). <i>Standards for educational and psychological testing</i>, pp. 63-72.</p>	

El segundo grupo remite a cuestiones de validez de las interpretaciones de puntajes de las pruebas en relación a los usos propuestos para la población objetivo evaluada. Se consideran estándares relacionados con la manera de analizar las diferencias entre subgrupos de la población evaluada el “contenido de la prueba, estructura interna de las respuestas, relación de los puntajes con otras variables, y procesos de respuesta de algunos individuos... estudios sobre la precisión y exactitud de los puntajes” (AERA *et al.*, 2014, p. 65), la validez de criterio cuando se hagan predicciones basadas en los puntajes de las pruebas, y la validez de la interpretación de los puntajes, en especial cuando se consideran preguntas de respuesta construida.

El tercer grupo de estándares refiere a las acomodaciones para remover las barreras irrelevantes y apoyar las interpretaciones válidas de los puntajes y sus usos propuestos. Este grupo tiene que ver con los ajustes que se hacen para remover las barreras que son irrelevantes para la medición del constructo. Las adaptaciones pueden darse en cuanto al “lugar de la aplicación, presentación de la

prueba, interfaz, requerimientos de respuesta, o añadir otros individuos al proceso de aplicación” (AERA *et al.*, 2014, p. 67). También comprende la documentación de los procedimientos para definir las adaptaciones, la difusión de la información a los sustentantes, y el análisis de la validez de las interpretaciones, entre otros.

El cuarto grupo alude a garantías en contra de interpretaciones inadecuadas de los puntajes de la prueba con respecto a los usos pretendidos. Comprende cuatro estándares relacionados con la provisión de información que oriente la interpretación adecuada de los puntajes para los diferentes grupos de población que participaron en la prueba. Se pide que los desarrolladores de las pruebas provean evidencia de la aplicabilidad de la prueba para los diferentes grupos y que hagan explícitas las precauciones que se deben tomar para evitar usos inadecuados de los resultados (AERA *et al.*, 2014). También se requiere que cuando los:

resultados de investigaciones muestren que los puntajes de algunos subgrupos poblacionales fueron afectados diferencialmente por características no relacionadas con el constructo medido, ... sólo se usen los resultados de los subgrupos para los cuales hay suficiente evidencia de validez para apoyar las interpretaciones de puntaje para los usos propuestos (AERA *et al.*, 2014, p. 70).

En relación a la publicación de resultados, se señala que antes de agregar puntajes de diferentes subgrupos poblacionales se debe asegurar y proveer evidencia de su comparabilidad y en caso contrario advertir que los resultados no

son comparables. Se señala también que los resultados de las pruebas no pueden utilizarse como único elemento para tomar decisiones con respecto a la colocación de alumnos en programas de educación especial; se deben utilizar múltiples fuentes de información.

#### *Estándares del ETS para cuidar la equidad en las evaluaciones*

El ETS reconoce que la equidad en las evaluaciones es un aspecto transversal y que es un concepto muy relacionado con el proceso de validación, por ello estableció siete estándares para cuidar que las inferencias hechas a partir de los puntajes de una prueba sean válidas para los diferentes grupos de sustentantes, los cuales se presentan en la Tabla 5.

Asimismo, el ETS desarrolló guías específicas para promover la justicia de las evaluaciones respecto a tres fuentes de varianza irrelevantes al constructo que se pretende medir: cognitiva, afectiva y física (ETS, 2009). Las fuentes *cognitivas* de varianza irrelevantes del constructo medido ocurren cuando para responder un ítem se requieren habilidades o conocimientos no relacionados con el constructo que se desea medir. Las fuentes *afectivas* se presentan cuando se utiliza un lenguaje o imágenes que provoquen emociones fuertes por parte de los sustentantes que interfieran con su habilidad para responder un ítem correctamente. Y las fuentes *físicas* de varianza irrelevante ocurren cuando las características de la prueba interfieren con la habilidad de los sustentantes para asistir a la aplicación, ver, oír o percibir los ítems o estímulos de la prueba (ETS, 2009).

**Tabla 5. Estándares sobre equidad del Educational Testing Service, 2014.**

Núm. Estándar	Descripción del Estándar
5.1: Tratamiento de la equidad	Diseñar, desarrollar, administrar y calificar pruebas para que midan el constructo deseado y minimizar los efectos de las características irrelevantes al constructo de los examinados. Para un servicio o producto nuevo, o revisado de manera importante, proveer un plan para hacer frente a la equidad en su diseño, desarrollo, administración y uso. Para un programa en curso, documentar lo que se hizo en el pasado para abordar la equidad, así como documentar cualquier plan de equidad a futuro.
5.2: Revisar y evaluar la equidad	Obtener y documentar evaluaciones de inequidad de productos o servicios, realizadas a través juicios o, de ser posible, de forma empírica con los grupos estudiados. En su caso, representar a varios grupos en los materiales de la prueba. Seguir las directrices diseñadas para eliminar los símbolos, el lenguaje y el contenido que es generalmente considerado como sexista, racista u ofensivo, excepto cuando sea necesario para cumplir con el propósito del producto o servicio.
5.3: Proporcionar acceso justo	Proporcionar acceso imparcial a los productos o servicios. Para las pruebas, proveer registro, administración y reporte imparciales de los resultados.
5.4: Elección de las medidas	Cuando un constructo se puede medir de diferentes maneras que son razonablemente igual de válidos, confiables, prácticos y asequibles, considerar la evidencia disponible de las diferencias en los puntajes de los subgrupos para determinar cómo medir el constructo.
5.5: Provisión de acomodaciones y modificaciones	Proporcionar acomodaciones o modificaciones apropiadas para las personas con discapacidades y para aquellos cuya lengua nativa no sea el inglés, de conformidad con las leyes aplicables, las políticas del ETS y del cliente.
5.6: Reporte de puntuaciones agregadas	Si las puntuaciones agregadas se reportan de forma separada para los grupos estudiados, evaluar la comparabilidad de las puntuaciones de esos grupos con respecto a las puntuaciones de toda la población.
5.7: Tratamiento de las necesidades de hablantes no nativos del inglés	En el desarrollo y uso de los productos o servicios, considere las necesidades de los hablantes no nativos del Inglés que pudieran derivarse de diferencias nos relevantes en el idioma o aspectos culturales. Para las pruebas, reducir las amenazas a la validez que podrían surgir de las diferencias del idioma y aspectos culturales.

**Fuente:** traducción propia a partir de ETS (2014). *ETS Standards for Quality and Fairness*, pp. 19–22.

Los estándares de la AERA-APA-NCME y del ETS buscan asegurar, en la medida de lo posible, que haya una atención y ocupación en aspectos de justicia de los procedimientos que se utilizan para diseñar, administrar, calificar y usar los resultados de las evaluaciones. En la medida en que los diseñadores de pruebas cumplan con ellos, las decisiones que se tomen a partir de los resultados de las evaluaciones serán mucho más justas y basadas en una medición más objetiva.



### **2.2.2. Aspectos de validez diferenciada**

En términos técnicos, el sesgo se concibe como “validez diferenciada de la interpretación de los puntajes de una prueba para cualquier subgrupo relevante y definido de examinados” (Cole y Moss, 1989, p. 205). Existe una estrecha vinculación entre equidad y validación en las evaluaciones; por eso, en la definición anterior el sesgo se considera como validez no uniforme, que difiere entre los grupos de participantes en la prueba, lo que se conoce como validez diferenciada.

En este apartado se presentan cinco categorías bajo las cuales se puede investigar la validez diferenciada o el sesgo. No se trata de cinco tipos distintos de sesgo, sino de cinco facetas bajo las cuales se le puede estudiar. Al igual que ocurre con los estudios de validez en los que se analizan cinco fuentes de evidencias, y que todas hacen alusión al concepto unitario de validez, en el análisis del sesgo se concibe algo similar. Las cinco categorías son: 1) contextualización del constructo; 2) contenido y formato; 3) administración y calificación de la prueba; 4) estructura interna de la prueba; y 5) relaciones externas del test (Cole y Moss, 1989)

#### *Contextualización del constructo*

En las investigaciones sobre validez y sesgo, el primer paso consiste en definir el uso que se le pretende dar a un test y cómo se ajusta el constructo que se pretende medir (Cole y Zieky, 2001). El mismo constructo puede asumir naturalezas distintas dependiendo del uso de la prueba. Por ejemplo, si el test

pretende utilizarse con propósitos de diagnóstico para realizar la planeación de las clases de matemáticas, el constructo puede incluir aspectos muy específicos como la habilidad para sumar fracciones mixtas, e incluir ítems que permitan identificar las causas de los errores cuando los alumnos no son capaces de encontrar un denominador común. En cambio, si el propósito del test es seleccionar a los estudiantes que pueden cursar cierta asignatura, el constructo puede ser definido de forma mucho más amplia, en términos de habilidades matemáticas, y la información que se requiere de los ítems es principalmente para tomar una decisión de selección y, por tanto, no es importante identificar el tipo de errores que cometen los sustentantes. Es decir, el mismo constructo, puede servir para múltiples propósitos pero generalmente no sirve para todos ellos igual de bien (Cole y Moss, 1989).

La consideración del contexto conlleva al establecimiento de la población de examinados que presentarán la prueba. Aquí, conviene especificar las poblaciones relevantes sobre las que se plantearán las hipótesis de sesgo, que luego se pondrán a prueba. En un contexto de recursos finitos para actividades de validación, conviene seleccionar las preguntas clave sobre validación que se harán para cada grupo relevante de examinados. Cuando los usos propuestos de las pruebas pueden derivar en decisiones importantes en los individuos o los sistemas (selección, colocación, acreditación, promoción, etc.), los análisis de sesgo se vuelven especialmente importantes (Cole y Moss, 1989).

Los estudios de contextualización del constructo buscan identificar el grado de ajuste entre: la especificación del constructo, las características del test y los usos

que se le pretenden dar a los resultados de la prueba. Para ello, los que realizan los estudios de validación, establecen hipótesis de trabajo que ponen a prueba a partir de las evidencias empíricas que sean recolectadas y de las evidencias lógicas que sean construidas. Básicamente, las hipótesis o preguntas de investigación que se plantean en el marco del proceso de validación son similares a las que se plantean en un análisis de sesgo, la diferencia con las últimas estriba en que el interés radica en indagar si la validez ocurre con cada grupo de individuos sobre los que se harán juicios de valor. Así, en los estudios de sesgo se plantean preguntas como: ¿las interpretaciones podrían ser menos válidas para ciertos grupos tales como ciertas etnias, mujeres, personas con alguna discapacidad, grupos cuya lengua familiar es distinta a aquella en la que está hecha la prueba, u otros? ¿la especificación del constructo implica algunos conocimientos o habilidades que son irrelevantes para los usos e interpretaciones que se le pretende dar los resultados de la prueba? ¿la especificación del constructo es apropiada para todos los grupos que presentarán la prueba? (Cole y Moss, 1989).

En resumen, los estudios sobre validez diferenciada que se realizan para encontrar evidencias de la contextualización del constructo, por lo regular se realizan a través de jueces y análisis lógico en el que el objetivo fundamental es poner a prueba la hipótesis de la coherencia entre la definición del constructo, las características del test y los usos que se le pretender dar a sus resultados, pero dicho análisis se lleva a cabo para cada una de las subpoblaciones relevantes de sustentantes.

### *Contenido y formato*

El contenido y formato de una prueba dependen del constructo que se desea medir y el contexto en que se usarán sus resultados. Los estudios de validez diferenciada que abordan la categoría de contenido y formato, suelen responder a esta pregunta, general: ¿El contenido y el formato son apropiados para el constructo que se pretende medir y los usos que se intentan dar a los puntajes, para cada uno de los grupos correspondientes? (Cole y Moss, 1989). Nótese que la primera oración refiere a un asunto meramente de validación y, es el segundo enunciado el que la vuelve una pregunta que atañe a la equidad en las evaluaciones, y por eso se considera como una validez diferenciada.

Uno de los métodos que se suelen utilizar para investigar el sesgo en el contenido y formato de una prueba consiste en incluir jueces que sean representativos de los grupos minoritarios sobre los que se quiere cuidar el sesgo. El trabajo de estos jueces consistirá en emitir juicios durante el proceso de desarrollo del test; estos jueces tendrán que revisar las especificaciones de los ítems, los lineamientos para los diseñadores de ítems, los reactivos mismos, las instrucciones de la prueba y el formato en que se presenta. Su trabajo esencialmente tendrá que consistir en identificar si el grupo al que representan tendría dificultades para responder la prueba ocasionadas por cuestiones de formato o de contenido que son ajenos a lo que se pretende medir, tales como la presencia de ciertos estereotipos, la representación de cierto grupo en el contenido de los ítems y su familiaridad con el contenido y el formato de la prueba.

Para ciertos grupos de sustentantes, el formato en el diseño de una prueba puede impedir que demuestren a cabalidad el estado del constructo que se desea medir. Por ello es importante que las evaluaciones cuiden la *accesibilidad* de los evaluandos, entendida como la oportunidad –sin obstáculos– que “debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir” (AERA-APA-NCME, 2014, p. 49). Por ejemplo, cuando el diseño tipográfico no permite que los sustentantes con alguna debilidad visual puedan leer los ítems de una prueba que no está destinada a medir la capacidad visual.

En los estudios de sesgo, es importante distinguir el sesgo aparente (*facial bias*) del sesgo real. El sesgo aparente ocurre cuando cierta palabra o formato en el ítem parece que afecta a cierto grupo, pero no altera las puntuaciones en el test. Por ejemplo, si una prueba de matemáticas para niños de secundaria contiene ítems cuyos problemas sólo emplean nombres de niñas, para una mirada hipercrítica podría dar la ilusión de que el test tiene un sesgo a favor de las mujeres; sin embargo, si esa situación no impide que los hombres realicen las tareas evaluativas, no se podría considerar un sesgo. Entre los psicómetras no existe una postura predominante sobre esta situación, sin embargo, es importante hacer patente que no todas las situaciones que parecen favorecer a un subgrupo son consideradas sesgo, sino sólo aquellas que impidan que esos grupos utilicen todos sus talentos para mostrar el grado en que poseen el constructo que se pretende medir.

### *Administración y calificación de la prueba*

La tercera categoría en el análisis de la validez diferencia alude a la administración y calificación de la prueba. Los protocolos de administración de un test tienen, en teoría, que permitir que todos los grupos demuestren su máximo potencial, lo cual requiere ciertas condiciones físicas, instrucciones claras sobre cómo responder la prueba, que el aplicador realice con los sustentantes una adecuada motivación para enfrentarse al test y un tiempo adecuado para contestarlo (Cole y Moss, 1989).

La estandarización en la aplicación de pruebas ayuda a que las condiciones y los procesos en que se administran los test sean en general apropiados; sin embargo, cualquiera de las condiciones de administración (físicas, en la instrucción, motivación o tiempo) que afecte diferenciadamente a los grupos puede producir sesgo. Hay unas cuestiones más evidentes que otras; así, resulta claro que en una prueba con un límite de tiempo establecido, si al momento de su aplicación, en uno de los grupos se otorga más o menos tiempo del establecido, ese grupo se verá favorecido o afectado, por lo que la validez de los resultados tendría que ponerse en duda; sin embargo, hay otras cuestiones que son más sutiles, como la motivación que propicia el aplicador de la prueba, sobre todo en test de bajo impacto, donde no habrá consecuencias directas para los alumnos.

En los estudios enfocados a identificar sesgo en la administración de pruebas, habrá que ser muy cuidadosos en considerar las condiciones y contexto de los sustentantes, y evitar el sesgo aparente. Por ejemplo, a raíz de los movimientos

de atención a la diversidad, los alumnos con alguna discapacidad son integrados a la escuela regular al tiempo que reciben educación especial en dichos centros, en consecuencia, las pruebas educativas intentan adaptarse a dichas poblaciones e incluir ciertas acomodaciones. Para un investigador del sesgo poco entrenado, el hecho de que mientras responden a una prueba de matemáticas el aplicador les permita a ciertos estudiantes utilizar una calculadora y a otros no, podría ser un indicador de invalidez en la administración; sin embargo, cuando se considera que eso es una acomodación para alumnos que padecen cierto retraso mental, esa misma evidencia se podría considerar como algo positivo y que atiende a la equidad en las evaluaciones.

El sesgo en la calificación de la prueba ocurre, sobre todo, en ítems de respuesta abierta donde se requieren jueces para su calificación. Las preocupaciones sobre el acuerdo inter-jueces (*inter-rater agreement*) se vuelven preocupaciones de sesgo cuando los calificadores se ven influenciados por las características de los grupos evaluados para emitir una calificación, y que dichas características son irrelevantes al constructo y propósitos de la prueba (Cole y Zieky, 2001). Las expectativas (conscientes o inconscientes) del calificador sobre las habilidades de los evaluados a partir de su origen social, escuela de procedencia, sexo, etc. así como la falta de criterios explícitos y detallados sobre lo que se considerará una respuesta correcta, son algunos de los elementos que podrían sesgar los resultados de un grupo de alumnos en particular.

Para indagar el sesgo en el proceso de calificación, los estudios suelen incluir una revisión de las rúbricas que emplean los calificadores, el proceso de capacitación

y perfil de los calificadores, así como estudios sobre el grado de coincidencia entre jueces. Sin embargo, los estudios a partir de la Teoría de la Generalizabilidad (Shavelson y Webb, 1991, 2006) ofrecen un instrumental estadísticamente más sólido para identificar facetas de error que dependiendo del diseño de la investigación, podrían aducirse a sesgo durante el proceso de calificación.

### *Estructura interna de la prueba*

Los estudios de validez sobre la estructura interna de las interpretaciones y usos de pruebas se centran en identificar la interrelación que existe entre las partes de un test (AERA-APA-NCME, 2014). Esas partes pueden ser ítems individuales o bloques de ítems. Los estudios de sesgo que se realizan bajo esta cuarta categoría, buscan identificar si las interrelaciones entre las partes del test permanecen en todos los grupos relevantes de sustentantes (Cole y Moss, 1989).

En los análisis que se basan sólo en identificar las relaciones entre ítems o puntajes sin que haya un criterio externo, como son los análisis de estructura interna, no se pueden detectar sesgo cuando todos los ítems o puntajes son influenciados por el elemento que los sesga. Estos análisis sólo permiten identificar diferencias en la estructura entre partes del test que son anómalas en algún grupo en comparación con la estructura de interrelaciones con el resto de grupos. Como señalan Cole y Moss (1989, p. 208): “Esa limitación ilustra la importancia de visualizar al sesgo como un problema integral que involucra varios tipos de evidencias más que un solo tipo, como los análisis de estructura interna”.



Existen múltiples procedimientos para analizar la estructura interna. Aquéllos que utilizan el análisis factorial, los que comparan la dificultad de los ítems, los que comparan las curvas características de los ítems (CCI) y los que utilizan aproximaciones a las CCI.

Los tipos de método que emplean los análisis factoriales, suelen comparan la estructura factorial de toda la población con la estructura de un grupo en particular; o bien, hacer comparaciones entre las estructuras factoriales de grupos de sustentantes en específico. A pesar de que hay un desarrollo importante al respecto a partir del modelamiento de ecuaciones estructurales (Brown, 2015; Lei y Wu, 2007; Reise, Widaman y Pugh, 1993), una de las principales dificultades es que a pesar de que se dispone de estadígrafos que miden la disparidad entre las estructuras en comparación, no se ha logrado consolidar hasta qué punto se puede sospechar que exista sesgo para alguna de las subpoblaciones estudiadas.

El segundo de los tipos de métodos, se basan en la comparación de las dificultades de los ítems entre grupos de interés. La dificultad de los ítems, en este tipo de métodos consiste en calcular la proporción de aciertos o una transformación que vuelva lineal dicha variable. El trabajo de Angoff es ilustrativo de este tipo de metodologías; donde a partir de variable transformada de la proporción de aciertos generaba una gráfica de dispersión, donde en el eje de las X representaba a uno de los grupos en comparación y en el eje de las Y al otro grupo de interés, y los puntos en la gráfica que representaban el nivel de dificultad de cada reactivo permitían visualizar la correlación entre los niveles de dificultad de ambos grupos. El problema con estos métodos es que no consideran el nivel

de habilidad de los sustentantes ni el nivel de discriminación de los ítems. Así, aun cuando en la gráfica la correlación pareciera muy fuerte, eso no necesariamente indicaría la ausencia de sesgo, puesto que las comparaciones ignoran el nivel de habilidad de los sustentantes (Shepard, Camilli y Williams, 1985).

En el tercer tipo de métodos para analizar la estructura interna de las pruebas, se comparan las CCI de los grupos bajo análisis. Las CCI se generan considerando el nivel de habilidad de los sustentantes (en el eje X) y la probabilidad de responder de manera correcta al ítem (en el eje Y). Un reactivo sin sospecha de sesgo sería aquel en el que su CCI fuera muy similar en cada uno de los grupos de sustentantes; en cambio, un ítem o un grupo de ellos se marcaría con amenaza de sesgo cuando sus CCI son distintas con al menos uno de los grupos de interés (Camilli, 2006).

El cuarto tipo de métodos son aproximaciones a la comparación entre CCI, donde el nivel de habilidad de los sujetos se genera a partir de una aproximación, creando intervalos a partir del puntaje total en la prueba; para cada reactivo dentro de cada intervalo del nivel de habilidad se tabula la frecuencia de respuestas correctas e incorrectas. Estos métodos pueden denominarse como pseudo TRI, que resultan muy apropiados cuando alguno de los grupos de interés es reducido, y no se cubre el requisito de muestras grandes que se requieren para los métodos de la tercera categoría (Shepard *et al.*, 1985).

*Relaciones externas del test*

Según Cole y Moss, “Las interpretaciones que se le dan al puntaje en una prueba implican numerosas formas de relación entre el puntaje y otras variables externas al test” (1989, p. 212). El análisis de sesgo, bajo esta perspectiva, consiste en identificar si esas relaciones son diferentes para ciertos grupos, lo que significaría una validez diferenciada.

Las variables externas pueden ser aquéllas que tienen un sustento teórico, o bien, con una variable que predice, después de un tiempo de haber sido aplicado el test. En todo caso, lo importante es identificar si la estructura de relaciones entre los puntajes de la prueba con las otras variables permanece o es sustancialmente diferente.

Los estudios que relacionan los puntajes de una prueba con variables externas, pretenden aportar evidencia convergente y discriminante. Se dispone de evidencia concurrente cuando los resultados de un test están altamente correlacionados con una variable externa que la teoría señala que lo está, o bien, con una variable externa que es conceptualmente similar. Por su parte, se acumula evidencia discriminante, cuando los puntajes de una prueba tienen una baja correlación con variables externas que la teoría las identifica en ese sentido o que conceptualmente son distintas al constructo que se pretende medir (Moss, 1992).

Para concluir este subapartado, vale la pena advertir que si bien hay varios procedimientos para identificar el sesgo en las mediciones educativas, el DIF es una de las herramientas más empleadas con este propósito (McNamara y Roever,

2006). Si bien existen múltiples técnicas para la detección de DIF, todas ellas tienen un elemento común: verificar que la probabilidad de respuesta correcta/incorrecta a cada ítem en un examen, sea independiente de la pertenencia de los respondientes a grupos de clasificación definidos por variables socio-demográficas y homologados por sus niveles de habilidad medida (González-Montesinos y Jornet, 2012).

El DIF es una condición necesaria pero no suficiente para que exista sesgo a favor o en contra de grupos de respondientes. El sesgo en ítems se presenta cuando además de la presencia confirmada de DIF, el funcionamiento diferencial se puede atribuir a características conceptuales en la construcción de reactivos que ponen en ventaja o desventaja la capacidad de emitir la respuesta correcta debido a que los respondientes pertenecen a un grupo determinado, y no al *impacto* de los ítems, que ocurre cuando las diferencias entre grupos de poblaciones se origina en las diferencias reales en los niveles de habilidad de dichos grupos.

Por su parte, hay autores que recomiendan la corroboración de sesgo en los ítems a partir de entrevistas cognitivas (Benítez y Padilla, 2014; Castillo-Díaz y Padilla, 2013; Ercikan *et al.*, 2010). Las entrevistas cognitivas son un método general que se puede usar para evaluar el grado en que la información que se pretende medir mediante ciertos instrumentos es coincidente con lo que los sujetos deseaban proporcionar. En particular, se usan para estudiar la manera en que la población objetivo a la que se dirigen los instrumentos son capaces de entender, procesar mentalmente y responder a dichos instrumentos, con especial énfasis en potenciales desajustes en el proceso (Willis, 2005, p. 3). Las entrevistas cognitivas

son uno de los métodos más robustos para conseguir evidencias de validez sobre los procesos de respuesta (Padilla y Benítez, 2014). Estos rubros se desarrollarán en apartados posteriores del marco teórico.

### **2.3. Funcionamiento Diferencial de los Ítems**

En términos técnicos, el DIF se presenta cuando existen “diferentes probabilidades para responder correctamente a un ítem por parte de sujetos con el mismo nivel de la característica medida, dependiendo del grupo al que pertenezcan” (Hidalgo y Gómez-Benito, 2010, p. 36). Gipps y Stobart señalan que “el DIF ocurre cuando personas en diferentes grupos se desempeñan en las preguntas de una prueba de maneras sustancialmente diferentes, incluso pudiendo tener puntajes muy similares” (2010, p. 56). Es decir, el DIF se presenta cuando grupos de estudiantes con el mismo nivel de habilidad, por el hecho de pertenecer a determinado grupo, tienen diferentes probabilidades de responder correctamente a un reactivo.

Al realizar estudios de DIF es importante distinguir entre impacto y funcionamiento diferencial. El impacto se refiere a “las diferencias observadas en el desempeño de un ítem por dos poblaciones diferentes que son diferencias reales en la habilidad de interés que es medida por la prueba” (Hidalgo y Gómez-Benito, 2010, p. 36). Y por el contrario, el DIF hace alusión a diferencias en el desempeño en un ítem que no son reales, sino que están asociadas a la pertenencia a determinado grupo poblacional de los sustentantes.

En los estudios de DIF lo habitual es hacer una comparación entre dos grupos: el grupo focal y el grupo de referencia. El grupo focal es donde se sospecha que hay DIF y generalmente se identifica con este grupo a las poblaciones minoritarias o en desventaja (por ejemplo, grupos indígenas, de contexto socioeconómico bajo, con lengua materna diferente al que está hecha la prueba, etc.). El grupo de referencia es, generalmente, el grupo mayoritario. Sin embargo, algunas técnicas de DIF permiten hacer comparaciones entre múltiples grupos.

En términos generales, existen dos tipos de DIF, el uniforme y el no uniforme (NUDIF). El DIF se denomina uniforme cuando la probabilidad de responder correctamente a un ítem es mayor para uno de los grupos en todos los niveles de habilidad. Nótese que en la sección izquierda de la Figura 1, en todo el recorrido del eje de las X (que representa la habilidad medida) el grupo focal tiene más probabilidades de responder correctamente el ítem. Por su parte, el NUDIF ocurre cuando la probabilidad de responder correctamente a un ítem por parte de los grupos focal y de referencia no es igual en todos los niveles de habilidad de los sustentantes, sino que hay una interacción donde, por ejemplo, en la parte baja de la habilidad medida un grupo tiene más probabilidades de responder correctamente a la pregunta y en la parte alta de la habilidad otro grupo tiene más probabilidades de contestar correctamente al reactivo (véase la parte derecha de la Figura 1).



**Figura 1. Ejemplo de DIF uniforme (izquierda) y DIF no uniforme (derecha) en las Curvas Características de los Ítems**

Desde que a mediados del siglo XX se inició a estudiar el sesgo y hasta finales del siglo XX se crearon muchas técnicas para analizar el DIF y algunas formas de clasificarlos (Camilli y Shepard, 1994). Estas autoras clasificaron las técnicas en tres grupos: los que se basan en la Teoría Clásica de los Test y métodos del análisis de varianza; los métodos basados en la Teoría de Respuesta al Ítem (TRI); y métodos a partir de tablas de contingencia. No obstante, derivado de los avances en el desarrollo de métodos para la detección de DIF, surgieron otras clasificaciones más recientes.

La clasificación de Hidalgo y Gómez-Benito (2010) es la que se adopta en este texto. Ellas agrupan las técnicas de DIF a partir de tres criterios: el tipo de modelo (paramétrico o no paramétrico), el tipo de variable de correspondencia (observada o latente) y el tipo de ítem (binario o politómico). Esta clasificación se presenta en la Tabla 6.

**Tabla 6. Clasificación de las técnicas de DIF**

Formato del ítem	Variable de correspondencia	Modelo	Técnica DIF
Binario	Puntaje observado	No paramétrico	Mantel–Haenszel (MH) Estandarización
		Paramétrico	Regresión logística Modelos Log-lineales Modelos de Clase Latente
	Variable latente	No paramétrico	SIBTEST CATSIB
		Paramétrico	Cociente del logaritmo de verosimilitud TRI Ji-cuadrada de Lord Mediciones del área Prueba del Funcionamiento Diferencial del Ítem en TRI Análisis Factorial Confirmatorio Indicador múltiple-Causa múltiple
Politómica	Puntaje observado	No paramétrico	MH generalizado MH ordinal Estandarización
		Paramétrico	Regresión logística multinomial Regresión logística discriminante Modelos Log-lineales Modelos de Clase Latente
	Variable latente	No paramétrico	POLYSIBTEST
		Paramétrico	Cociente del logaritmo de verosimilitud TRI Ji-cuadrada de Lord Mediciones del área Prueba del Funcionamiento Diferencial del Ítem en TRI Análisis Factorial Confirmatorio Indicador múltiple-Causa múltiple

Fuente: Hidalgo y Gómez-Benito, 2010, p. 38.

### 2.3.1. Técnicas no paramétricas para detección del DIF

Las técnicas estadísticas que se agrupan en esta clasificación por lo regular se basan en tablas de contingencia y se pueden distinguir tres enfoques en su interior: 1) aquellas técnicas que involucran pruebas de hipótesis sobre la igualdad de las proporciones tales como la Ji cuadrada, el procedimiento de



estandarización y el estadígrafo de Mantel-Haenszel (M-H); 2) técnicas que usan un modelo categorial tales como los modelos log-lineales, modelos de clase latente y análisis de regresión logística; y 3) el procedimiento de prueba simultánea de sesgo en los ítems (SIBTEST, por sus siglas en inglés). Dentro de cada enfoque, las técnicas más empleadas son el M-H, el análisis de regresión logística y el SIBTEST, respectivamente (Hidalgo y Gómez-Benito, 2010).

La técnica de M-H fue utilizada por primera vez para analizar el DIF por Holland y Thayer (1988). En la actualidad esta técnica es una de las más empleadas y conocidas porque el ETS, la agencia más grande diseñadora de pruebas en EU, la emplea para analizar el DIF en los ítems de sus test. El procedimiento M-H consiste en generar tablas de 2X2 en  $k$  grupos de habilidad de los sustentantes. Por lo regular se divide a los evaluados en 5 o 6 grupos o intervalos ( $k$ ) según el porcentaje de preguntas que respondieron correctamente. En cada grupo se calcula el momio o posibilidad de respuestas correctas y se compara este valor entre el grupo focal y el de referencia.

Para medir la magnitud del DIF, el procedimiento original de M-H se basaba en el cociente de posibilidades común ( $\alpha$ ); sin embargo, Holland y Thayer (1988) propusieron una transformación logarítmica de ( $\alpha$ ) para mejorar su interpretación; esta transformación se expresa como  $\Delta_{\alpha MH} = -2.35 \ln(\alpha)$  lo que sirve para tener una escala simétrica en la que los valores cercanos a cero son indicativos de ausencia de DIF, valores negativos significan que el grupo de referencia se ve favorecido mientras que valores positivos indican que el grupo focal tiene mayores posibilidades de responder a determinado ítem de forma correcta. Valores

absolutos menores a 1 en la escala delta del M-H ( $\Delta_{\alpha MH}$ ) representan ausencia de DIF o DIF insignificante; cuando  $\Delta_{\alpha MH}$  toma valores entre 1 y 1.5 el ítem presenta un DIF moderado; y cuando  $\Delta_{\alpha MH}$  es mayor que 1.5 se considera que el ítem tiene un DIF severo (Zwick, 2012).

Las principales ventajas del método M-H son su sencillez para calcularse, su gran efectividad para identificar DIF y la baja demanda computacional por lo que se incluye en software estadístico de uso general como SPSS, SAS y STATA, entre otros. Además, es relativamente fácil de interpretar por profesionales sin experiencia en psicometría y no requiere muestras muy grandes de sujetos, de 200 personas en cada grupo (de referencia y focal). Sin embargo, tiene al menos las siguientes limitaciones: a) no permite detectar NUDIF, aunque Clauser *et al.* (1994) propusieron una variación del método que resultó con buenos niveles de detección del NUDIF; b) pobre desempeño para detectar DIF con muestras menores a 100 sujetos; y c) se basa sólo en las puntuaciones observadas, cuando la variable que se pretende medir pertenece en realidad a un constructo o variable latente (Fidalgo, 1997; Hidalgo y Gómez-Benito, 2010).

La regresión logística es una técnica muy flexible que permite identificar diferentes tipos de DIF (uniforme y no uniforme), trabajar con distintos tipos de ítems (binarios y politómicos), utilizar distintos criterios de correspondencia y variables de agrupamiento con dos o más niveles, como género, raza, lengua, etc. (Balluerka, Gorostiaga, Gómez-Benito e Hidalgo, 2010; García, Gómez-Benito e Hidalgo, 2005; Hidalgo, Gómez-Benito y Zumbo, 2014; Svetina y Rutkowski, 2014; Zumbo, 1999). Como describe Fidalgo:

Mediante esta técnica se trata de determinar si en la función matemática necesaria para predecir las respuestas dicotómicas a un ítem se debe incluir un término referido a la interacción entre el grupo y la habilidad (DIF no uniforme), o al grupo de pertenencia (DIF uniforme), o simplemente puede predecirse en función del nivel de habilidad de los sujetos con independencia de su grupo de pertenencia (ausencia de DIF) (1997, pp. 410–411).

Unas de las ventajas de esta técnica son: a) ofrece una prueba estadística sobre la presencia de DIF y dispone de un estadígrafo sobre el tamaño del efecto; b) tiene buena potencia para detectar DIF con muestras relativamente pequeñas; y c) se puede implementar a través de software estadístico de uso general. Las limitaciones del análisis de regresión logística son: a) los requerimientos para su tamaño de muestra depende del número de categorías de respuesta y de su grado de asimetría, de tal manera que para ítems sin asimetría, tamaños de muestras más pequeños (por ejemplo 100) puede ser inadecuados para detectar tanto DIF uniforme como no uniforme; b) funciona con las puntuaciones observadas, y como ya se señaló con anterioridad, esto es una limitación porque el constructo que se pretende medir es una variable latente; y c) puede requerir un proceso de purificación de los puntajes de correspondencia con unos cuantos ítems (Hidalgo y Gómez-Benito, 2010).

La técnica denominada prueba simultánea de sesgo en los ítems (SIBTEST, por sus siglas en inglés) se basa en la teoría multidimensional del sesgo y en un modelo multidimensional no paramétrico dentro de la TRI, propuestos por Shealy y Stout (1993). Su principal ventaja frente a otras técnicas es que permite analizar la

presencia de DIF en varios ítems de manera simultánea lo que a su vez permite determinar si existe una amplificación o cancelación del DIF en el test, lo que da lugar al DFT (funcionamiento diferencial del test). La amplificación ocurre cuando los ítems con DIF actúan en conjunto para afectar las puntuaciones del test de alguno de los grupos de pertenencia (focal o de referencia), en cambio, la cancelación es el fenómeno opuesto, donde a pesar de la presencia de DIF en varios ítems, su efecto se anula porque ciertos ítems favorecen a uno de los grupos mientras que otros lo desfavorecen (Fidalgo, 1997).

A partir del SIBTEST, que originalmente se utilizaba para ítems dicotómicos y detección de DIF uniforme, se desarrollaron extensiones de este método para el trabajo con ítems politómicos (Poly-SIBTEST) y evaluar el NUDIF (Crossing-SIBTEST). Además, la extensión CATSIB en lugar de operar con los puntajes observados, emplea las puntuaciones estimadas basadas en la TRI. Las principales desventajas del SIBTEST son: a) que el criterio para determinar al subtest válido o que no tiene ítems con DIF requiere de un procedimiento muy complejo; b) no es muy potente para detectar DIF con muestras pequeñas; y c) la detección de NUDIF mediante Poly-SIBTEST se probó que es muy problemática (Hidalgo y Gómez-Benito, 2010).

### **2.3.2. Técnicas paramétricas para detección del DIF**

En el marco de la TRI, existe muchas técnicas para la detección de DIF, tales como la comparación de modelos a través de la razón de verosimilitud, la comparación de parámetros ( $\chi^2$  cuadrada de Lord), la medición de áreas dentro de las CCI y los métodos basados en el marco del funcionamiento diferencial de

ítems y pruebas (Hidalgo y Gómez-Benito, 2010). En TRI, los modelos de un parámetro sólo consideran el nivel de dificultad ( $\delta$ ) de los ítems, y es conocido como el modelo de Rasch; los modelos de 2 parámetros incorporan, además, el índice de discriminación ( $\alpha$ ); y en el modelo de 3 parámetros se agrega también el factor de adivinación ( $c$ ) o la probabilidad de responder a un reactivo cuando el sujeto tiene cero de la habilidad medida (Boone, Staver y Yale, 2014).

Las pruebas de razón de verosimilitud y la Ji cuadrada ( $X^2$ ) de Lord se basan en la comparación de los parámetros de los ítems que fueron obtenidos en grupos distintos. Los test de razón de verosimilitud implican la comparación de modelos en los que se supone que los parámetros de uno o más ítems varían entre el grupo focal y de referencia, y aunque existe varios estadígrafos para hacer dicha comparación, el que más se suele utilizar es la devianza ( $-2Ln$ ). La razón de verosimilitud que está dada por:

$$RV = -2Ln \frac{V(A)}{V(C)}, \quad (1)$$

donde  $Ln$  es el logaritmo natural neperiano,  $V(C)$  es la verosimilitud del modelo compacto y  $V(A)$  es la verosimilitud del modelo ampliado. El modelo ampliado incluye todos los parámetros del modelo compacto y otros más; el propósito es identificar si los parámetros adicionales del modelo ampliado son estadísticamente distintos de cero, lo que indicaría, de ser el caso, una mejora significativa del ajuste del modelo a los datos (Camilli, 2006).

Por su parte, con la técnica de Lord los parámetros se estiman por separado para cada grupo y luego son comparados de manera directa; dependiendo del número de parámetros del modelo de medida empleado, se compara el nivel de dificultad, el índice de discriminación y el factor de adivinación. En el modelo de Rasch, la  $X^2$  de Lord se obtiene como:

$$X^2 = \frac{(\hat{b}_F - \hat{b}_R)^2}{Var(\hat{b}_F) + Var(\hat{b}_R)}, \quad (2)$$

donde  $\hat{b}_F$  y  $\hat{b}_R$  son los niveles de dificultad estimadas del ítem  $i$  para el grupo de referencia y focal, respectivamente,  $Var(\hat{b}_F)$  y  $Var(\hat{b}_R)$  son las varianzas estimadas del parámetro  $b$  en cada grupo (Fidalgo, 1997).

Las medidas del área, aunque también estiman por separado los parámetros de los ítems en cada grupo, el método consiste en calcular el área entre las CCI del grupo de referencia y focal. Para calcular el área se propusieron varios índices, algunos disponen de signo (lo que permite saber la dirección del DIF) mientras que otros son en términos absolutos lo que sólo permiten saber la magnitud del área; otros índices crean intervalos infinitos o finitos en la habilidad de los sustentantes para calcular el área; en otros índices la aproximación es continua y se calcula por medio de integrales, mientras que otros hacen una aproximación discreta y el procedimiento de cálculo incluye la suma de las áreas en cada intervalo de la habilidad; además, hay otros índices que ponderan las diferencias entre las posibilidades considerando la densidad de población, mientras que otros no realizan dicha ponderación. En todos ellos, mientras más pequeña sea el área representan un menor DIF (Hidalgo y Gómez-Benito 2010).

Los procedimientos dentro de la TRI que se basan en el Funcionamiento Diferencial de los Ítems y los Test (DFIT, por sus siglas en inglés) operan a partir de las puntuaciones esperadas por el modelo de medida en cuestión, las que se consideran puntuaciones verdaderas. Estas técnicas estiman la magnitud de la diferencia entre la puntuación predicha por el modelo de un sujeto como miembro del grupo de referencia y la puntuación esperada del mismo sujeto como miembro del grupo focal. Como lo describen Hidalgo y Gómez-Benito “cuanto mayor sea la diferencia entre estas puntuaciones, mayor es el Funcionamiento Diferencial del Test” (2010, p. 40).

Las principales ventajas de todos estos métodos dentro de la TRI son que están basados en un modelo teórico bien fundamentado, permiten identificar DIF y NUDIF, y pueden medir el funcionamiento diferencial de ítems tanto dicotómicos como politómicos. Sin embargo, también tienen algunas limitaciones, por ejemplo, se requieren ciertos tamaños de muestra para tener estimaciones apropiadas de los parámetros, el modelo tiene que ajustarse a los datos, se deben conocer los supuestos que el modelo asume (unidimensionalidad e independencia local) y requieren software específico que por lo regular no es muy amigable (Boone *et al.*, 2014; Hidalgo y Gómez-Benito, 2010; Zumbo, 1999).

### **2.3.3. Innovaciones en el análisis DIF**

Desde la sexta década del siglo XX hasta nuestros días, ha habido un gran desarrollo en técnicas para la detección del DIF, que siguen evolucionando continuamente. Además de los procedimientos para el análisis del funcionamiento diferencial que ya se presentaron en el apartado anterior, existen otros que son

más novedosos y aunque no se dispone de la misma cantidad de estudios que demuestren su eficacia, su uso se fue extendiendo cada vez más. Entre los que destacan son los procedimientos a partir del modelamiento de ecuaciones estructurales (SEM) y el modelo lineal de test logístico (LLTM).

Dentro de SEM existen varios procedimientos para detectar DIF que se pueden agrupar en tres: análisis factorial restringido, análisis factorial confirmatorio multigrupo con media latente y estructura de covarianza (CFA-MACS) (Byrne, Shavelson y Muthén, 1989; Hernández y González-Romá, 2003; Ployhart y Oswald, 2004; Sharma, Durvasula y Ployhart, 2012) y el modelo de Indicador múltiple-Causa múltiple (MIMIC, en inglés) (Chun, Stark, Kim y Chernyshenko, 2016; Lee, Bulut y Suh, 2016). En general, estos enfoques basados en SEM son capaces de detectar DIF en modelos multidimensionales y proporcionar diversos índices de ajuste tanto en modelos de nivel local como globales. Una de sus grandes ventajas es que estos procedimientos no sólo permiten la evaluación del DIF, sino que también son útiles para evaluar la equivalencia de constructo y el análisis de invariancia en mediciones interculturales (Hidalgo y Gómez-Benito, 2010).

Dentro del modelo lineal de test logístico (LLTM) desarrollado por Fisher, el funcionamiento diferencial de facetas (DFF, en inglés) es un procedimiento que se emplea para medir el DIF y que fue utilizado por primera vez con este propósito por Engelhard (1992). La faceta se refiere a un grupo de ítems que comparten una característica en común, por ejemplo, aquéllos que evalúan en su conjunto algún subdominio (Xie y Wilson, 2008).



En el modelo de Rasch, cuando un ítem presenta DIF la expresión que lo representa es:

$$\textit{logit} = \beta_n - \delta_i + Z_n \gamma_i, \quad (3)$$

donde  $\beta_n$  es el nivel de habilidad del sujeto,  $\delta_i$  es el nivel de dificultad del reactivo,  $Z_n$  es la variable que indica el grupo al que pertenece y  $\gamma_i$  representa el parámetro de DIF para un ítem  $i$ ; esto implica que  $Z_n \gamma_i$  puede considerarse como una interacción entre el ítem  $i$  y el grupo de membrecía  $Z$ . De forma análoga, en el modelo LLTM, el DFF se obtiene a partir de

$$\textit{logit} = \beta_n - \sum n_k Q_{ik} + Z_n \sum \gamma_k Q_{ik}, \quad (4)$$

en la que  $n_k$  es el parámetro de dificultad de la propiedad del ítem  $k$ ,  $Q_{ik}$  es el peso del ítem  $i$  en la propiedad del ítem  $k$  y el término  $\gamma_k$  es el parámetro DFF para la propiedad del ítem  $k$ . La idea fundamental es identificar el funcionamiento diferencial en bloques de ítems que juntos tratan de medir alguna faceta del constructo, lo que según los autores que lo han empleado, permite hacer interpretaciones mucho más apegadas al uso que se le dará a los puntajes (Xie y Wilson, 2008).

#### **2.4. Análisis de sesgo en las pruebas**

Para realizar análisis de sesgo en el campo de la evaluación educativa y psicológica, uno de los modelos que tomó especial importancia es el diseño universal de las evaluaciones (DUE). En este trabajo el DUE se empleó como

marco conceptual para analizar una buena parte de las evidencias recolectadas en campo a través de entrevistas cognitivas. Por ello, este subcapítulo está dividido en dos subapartados, en el primero se presenta el DUE y en el segundo se desarrolla el tópico de las entrevistas cognitivas.

#### **2.4.1. El diseño universal de las pruebas**

La equidad en las evaluaciones implica considerar la diversidad de contextos de quienes responderán los instrumentos de evaluación como una posible fuente de varianza irrelevante al constructo que se desea medir, de tal manera que los puntajes obtenidos en una prueba no se vean sesgados por aspectos ajenos al interés de la evaluación (como el contexto social, cultural, económico, etc.). En la literatura reciente hay consenso en aceptar que la atención a la diversidad en las evaluaciones debe ser atendida desde su diseño (AERA-APA-NCME, 2014; Gipps y Stobart, 2010; Solano-Flores, 2011).

Para lograrlo es importante considerar su *accesibilidad*, entendida como la oportunidad –sin obstáculos– que “debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir” (AERA-APA-NCME, 2014, p. 49). Como señalan Ruiz *et al.* (2015):

Se parte del supuesto que el diseño de una prueba puede impedir el acceso a demostrar el estado de un constructo, por ejemplo, cuando el diseño tipográfico no permite que los sustentantes con alguna debilidad visual puedan leer los ítems de una prueba que no está destinada a medir la capacidad visual (Ruiz, Pérez, Langford y García-Medina, 2015).

Uno de los enfoques que se impulsó para maximizar la accesibilidad de las evaluaciones es el *diseño universal*. El DUE es una corriente que surge en la arquitectura y que luego es trasladada a otros campos, como el medio ambiente, el cuidado de la salud, las artes y posteriormente a la educación, de manera importante en las evaluaciones del aprendizaje aplicadas de forma masiva. En el campo de las evaluaciones de gran escala, el diseño universal se refiere a “aquellas evaluaciones diseñadas y desarrolladas desde el inicio para favorecer la participación del mayor número de estudiantes, y que los resultados lleven a inferencias válidas acerca del desempeño de todos los estudiantes” (Thompson, Johnstone y Thurlow, 2002, p. 5), con lo que se promueva “el acceso a la mayor cantidad de sustentantes, sin importar su raza, origen étnico, género, estado socioeconómico, discapacidad, lengua o contexto cultural” (AERA-APA-NCME, 2014, p. 50).

Así, se busca dar respuesta a las necesidades de los sustentantes que no interfieren con la medición del constructo, pero que si no fueran tomadas en cuenta, les impediría el acceso a la medición o limitaría la validez de las interpretaciones de los resultados. Por ejemplo: la impresión magnificada de una prueba de Matemáticas para alumnos con algún tipo de debilidad visual (la medición del constructo no tiene relación con el ajuste en el diseño y por tanto no se verá afectada) (Cfr. Ruiz *et al.*, 2015).

El diseño universal requiere que los desarrolladores de las pruebas

tengan claridad sobre el constructo o constructos a ser medidos, incluyendo el objeto de la evaluación, el propósito con el que se usarán los puntajes, las

inferencias que se harán sobre los puntajes, y las características de los sustentantes y subgrupos de la población objetivo de la prueba que pueden influenciar el acceso (AERA-APA-NCME, 2014, p. 50).

El diseño universal se basa en siete principios: 1) uso equitativo; 2) flexibilidad en el uso; 3) uso simple e intuitivo; 4) información perceptible; 5) tolerancia al error; 6) esfuerzo físico bajo; y 7) tamaño y espacio para el acceso y el uso (Center for Universal Design, 1997). A partir de esos siete principios, Thompson *et al.* (2002) definieron los siete elementos clave que deberían cubrir las evaluaciones de gran escala: 1. Evaluación inclusiva de la población; 2. Constructos definidos de forma precisa; 3. Ítems accesibles y no sesgados; 4. Sensible a las adaptaciones; 5. Procedimientos e instrucciones simples, claros e intuitivos; 6. Máxima legibilidad y comprensibilidad (*Maximum readability and comprehensibility*); y 7) Máxima visibilidad (*Maximum legibility*).

*Evaluación inclusiva de la población.* La inclusión implica que en evaluaciones de gran escala todos los estudiantes puedan participar, sin importar sus habilidades cognitivas, sus contextos culturales o lingüísticos. Las evaluaciones requieren medir un amplio rango de habilidades y destrezas de los sustentantes, asegurándose que los estudiantes con diversas necesidades de aprendizaje reciben las oportunidades para demostrar su competencia sobre un mismo contenido.

*Constructos definidos de forma precisa.* La definición precisa de los constructos alude al principio básico de la validez, donde una función fundamental de las evaluaciones es que midan lo que pretenden medir. Así, las evaluaciones bajo el

diseño universal, remueven todas las barreras cognitivas, sensoriales, emocionales y físicas no relacionadas con el constructo que se desea medir. Esto se le conoce como varianza irrelevante al constructo.

*Ítems accesibles y no sesgados.* La accesibilidad se entiende como la oportunidad sin obstáculos que “debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir” (AERA-APA-NCME, 2014: 49). No sesgados implica que el reactivo esté libre de contenido que podría resultar ofensivo o penalizar a ciertos grupos de estudiantes debido a sus características personales como raza, género, etnicidad, “discapacidad” o estatus socioeconómico. Se parte del supuesto de que algunos rasgos de una prueba pueden dificultar que algunos sustentantes muestren el manejo que tienen de lo que se quiere medir, por ejemplo, cuando el diseño tipográfico no permite que los sustentantes con alguna debilidad visual puedan leer los ítems de una prueba que no está destinada a medir la capacidad visual.

*Sensible a las adaptaciones.* Aun cuando los ítems de las evaluaciones hechas bajo el diseño universal serán accesibles para la mayoría de los sustentantes, habrá algunos grupos que requerirán algunas adaptaciones en las pruebas o en su logística de aplicación a fin de que puedan ser incluidos en las evaluaciones. Las adaptaciones (*accommodations*) son ajustes o modificaciones en la manera en que una prueba es presentada o respondida, que generalmente se realizan para alumnos con alguna discapacidad o aquéllos que están aprendiendo la lengua en la que está hecha la prueba, por ejemplo, la conversión de un test en Braille para alumnos ciegos o con extrema debilidad visual, el aumento del tamaño

de letra y la ampliación de las figuras de los ítems como acomodación para alumnos con debilidad visual intermedia, o bien, la traducción de una prueba al náhuatl para aplicarse a niños que esa sea su lengua materna y apenas estén aprendiendo español.

*Procedimientos e instrucciones simples, claros e intuitivos.* Las instrucciones y los procedimientos sobre cómo responder los test e ítems deben ser fáciles de entender para los sustentantes, sin importar su experiencia, conocimiento, dominio del lenguaje o nivel de concentración. Instrucciones con un lenguaje complejo podrían invalidar los resultados de la prueba porque el sustentante no comprendió cómo debería responder, aun cuando tenía desarrollada la habilidad que el test pretendía evaluar. Por ejemplo, cuando en una sección del cuadernillo del examen se le pide que rellene el inciso de la respuesta correcta y en la siguiente sección se le solicita que coloque un círculo sobre la opción que seleccione. Por el contrario, cuando en un test de comprensión de lectora se le pide al estudiante que identifique el número de párrafo en el que se presenta la idea principal del texto y los párrafos está numerados, de tal manera que se eliminen las posibilidades de que el sustentante elija un párrafo distinto al que seleccionó como respuesta correcta.

*Máxima legibilidad y comprensibilidad (Maximum readability and comprehensibility).* Las evaluaciones bajo el diseño universal reducen la “complejidad verbal y organizacional de los ítems mientras preservan su contenido esencial (i.e. las habilidades y conceptos que intentan medir) (Hanson *et al.*, 1998, p.2 citado en Thompson *et al.*, 2002). Para ello, una de las estrategias que se

utilizan en pruebas que no pretenden medir el dominio de una lengua, como matemáticas o ciencias naturales, es redactar los ítems y las instrucciones de la prueba empleando un *lenguaje plano*. Un lenguaje plano es “un lenguaje basado en el texto que es sencillo, conciso, y usa palabras de uso cotidiano para transmitir su significado. La meta de las estrategias de edición del lenguaje plano es mejorar la comprensibilidad del texto escrito mientras se preserva la esencia del mensaje” (Hanson, *et al.*, 1998, p.2 citado en Thompson *et al.*, 2002).

*Máxima visibilidad (Maximum legibility)*. La visibilidad se refiere a la capacidad de que algo sea descifrado con facilidad. El término puede ser aplicado a un texto, a varios tipos de gráficas, tablas e ilustraciones así como al formato de respuesta. En el caso de los textos, la legibilidad implica que se cubran las siguientes características de manera adecuada: contraste, tamaño de letra, espaciado entre caracteres, interlineado, tipo de letra, justificación, longitud de la línea y espaciado en blanco. En el caso de las gráficas, tablas e ilustraciones, la legibilidad se refiere a que los símbolos en ellas tienen que ser altamente distinguibles, especialmente si las figuras están en blanco y negro; además, sus títulos deben estar al lado de ellas, entre varias otras cuestiones. En lo referente al formato de respuesta, la legibilidad implica que el tamaño de las hojas del cuadernillo, los óvalos a rellenar y la diagramación esté acordes a la población que responderá el test; un cuadernillo en tamaño media carta, con un tamaño de letra de 11, no será apropiado para alumnos de primaria con debilidad visual, por ejemplo.

En esta investigación, el DUE se empleó como marco para analizar e interpretar la información obtenida a partir de los protocolos verbales, tema que se desarrolla en el siguiente subapartado.

#### **2.4.2. Los protocolos verbales y reportes verbales**

Dentro de la teoría de la validación, en las últimas décadas del siglo XX se hizo cada vez más evidente que los procesos de respuesta eran el centro neurálgico en los errores de medición, y que debería de considerarse como una faceta de fuentes de error, lo que finalmente fue considerado en la versión de 1999 de los Estándares de la (AERA-NCME-APA, 1999). Desde entonces, los procesos de respuesta se consideran como una fuente de evidencia de la validez. El supuesto fundamental es que los procesos cognitivos del respondente conducen la respuesta que dan en el ítem, y un entendimiento de la cognición es central para el diseño de reactivos y la comprensión y reducción de fuentes de error en las respuestas (Willis, 2005). La obtención de evidencias de validez sobre los procesos de respuesta consiste en un análisis empírico y teórico del ajuste que existe entre los procesos cognitivos que emplean los examinados para responder a las tareas que se le solicitan en determinado instrumento y su correspondencia con los procesos que se asume deben utilizar cuando se indaga el constructo medido (AERA-APA-NCME, 2014; Miller, Willson, Chepp y Padilla, 2014).

Dentro del campo de la medición, existen varias técnicas para medir los procesos de respuesta que pueden clasificarse en dos categorías: 1. Aquéllas que acceden directamente a los procesos cognitivos, como los protocolos verbales, los reportes verbales (estas dos conocidas como técnicas de pensamiento en voz alta), las



entrevistas cognitivas y los grupos focales; y 2. Aquéllas que proveen indicadores indirectos y que requieren una inferencial adicional, como los estudios del sendero de la vista y aquéllos que registran el tiempo de latencia de respuesta (Padilla y Benítez, 2014); sin embargo, las que se fueron extendiendo más y se volvió la más sobresaliente dentro del diseño de instrumentos de evaluación en educación fueron las técnicas de pensamiento en voz alta, que incluyen a los protocolos verbales y los reportes verbales (Miller *et al.*, 2014).

Los protocolos verbales consisten en una serie de indicaciones para promover que el entrevistado, al tiempo que responde cada una de los ítems que se desea analizar, verbalice los procesos cognitivos que pone en juego para dar respuesta a las tareas evaluativas (Ericsson y Simon, 1993). La función principal de estos protocolos es que el sujeto se mantenga hablando desde que inicia a leer el pie del reactivo hasta que elige alguna de las opciones de respuesta (Leighton, 2009). El análisis de los protocolos verbales se suele utilizar para identificar y medir el proceso de resolución de problemas, especialmente para identificar los contenidos de la memoria de trabajo de los sustentantes (Ericsson y Simon, 1993), los cuales están asociados con las funciones ejecutivas de la memoria.

Los protocolos verbales incluyen una serie de pautas para que el examinado verbalice mientras evoca los conocimientos y procesos cognitivos con los que se enfrenta a los ítems. Los protocolos suelen incluir preguntas de sondeo (*probes*), que se realizan generalmente para obtener información sobre aspectos muy específicos de los ítems en los que se anticipa la existencia de ciertos problemas en su formulación. Las preguntas de sondeo que se incluyen en el protocolo son

redactadas para corroborar o descubrir si un reactivo contiene términos que pueden resultar confusos o son desconocidos para ciertos subgrupos poblaciones, una estructura sintáctica que dificulta la comprensión de la tarea, para profundizar sobre las operaciones cognitivas que los individuos pusieron en juego, constatar si para resolver determinado ítem los sujetos tuvieron que recordar determinados conocimientos o validar el modelo cognitivo del conjunto de reactivos en un test (Benítez y Padilla, 2014; Castillo-Díaz y Padilla, 2013).

Los reportes verbales, por su parte, consisten en elicitar que el sustentante vaya pensando en voz alta, pero a diferencia de los protocolos verbales, una vez que el examinado inicia a responder los ítems, no se le dan instrucciones adicionales y no se le interrumpe durante el proceso de resolución, salvo para solicitar que continúe verbalizando su pensamiento. Con este método, por lo general se suele indagar en la memoria de largo plazo de los examinados, la cual contiene los esquemas mentales más estables y con los cuales el entrevistado suele operar de forma cotidiana.

Cada una de las técnicas tiene ventajas y desventajas. Así, en los protocolos verbales, su principal ventaja es que permiten indagar cierto proceso cognitivo que no está siendo del todo explicitado por el examinado, al tiempo que piensa en voz alta mientras responde a un reactivo, y eso permite conocer con bastante detalle cada uno de los pasos que va siguiendo el informante; por el contrario, tiene la desventaja de que el proceso cognitivo "natural" se puede ver afectado, pues cuando el entrevistador le plantea preguntas de sondeo incluidas en el protocolo,

el entrevistado puede percatarse de algunas operaciones cognitivas que debió emplear y modificar la senda de su proceso cognitivo.

Por su parte, los reportes verbales, que tienen la ventaja de que el examinado haga una introspección en su memoria de largo plazo y utilice un proceso cognitivo natural; en cambio, la principal desventaja es que el entrevistado podría no ser capaz de recordar con precisión el proceso cognitivo que empleó y hacerlo saber o, en el peor de los casos, falsear la información (Willis, 2005, 2015).

Los estudios en los que se aplicaron protocolos verbales para explorar o confirmar fuentes de DIF en los ítems (Ercikan *et al.*, 2010), emplearon una combinación de técnicas de pensamiento en voz alta, protocolos y reportes verbales. En el estudio de Ercikan *et al.* (2010), al principio pidieron que los entrevistados pensaran en voz alta mientras respondían cada uno de los ítems identificados con DIF (reportes verbales), y luego, cuando el entrevistado seleccionaba alguna de las opciones de respuesta, utilizaron un protocolo verbal donde al examinado le plantearon preguntas de sondeo para conocer si tuvieron dificultad en algunos términos, la manera en que comprendieron lo que tenían que realizar, qué aspectos del reactivo le ayudaron a resolverlo (si es que los hubo) y cómo llegó a seleccionar la respuesta elegida. Con esta estrategia, se evita alterar el proceso que sigue el entrevistado para responder el reactivo, y dependiendo del tipo de preguntas de sondeo que se planteen, se puede tener una visión relativamente completa de las operaciones cognitivas que emplea así como los aspectos irrelevantes al constructo que intervinieron para que el sujeto respondiera de manera correcta o incorrecta.

### 3. MÉTODO

En este tercer capítulo se presenta con detalle el diseño de la investigación, los participantes, los instrumentos así como los procedimientos empleados. El estudio buscó aportar, desde un enfoque de equidad de las pruebas, evidencias de validez de los Exámenes de la Calidad y el Logro Educativos (Excale) para tercero de secundaria en Matemáticas<sup>4</sup>, mediante análisis del funcionamiento diferencial de ítems (DIF) y protocolos verbales. En específico, se buscó identificar si existe algún sesgo de la prueba para las subpoblaciones<sup>5</sup> según sexo y nivel socioeconómico (NSE) que se convirtieron en ejes de análisis transversales durante todas las etapas del estudio.

El estudio se desarrolló en dos etapas. La primera donde se hicieron los análisis DIF, se desarrolló en cuatro fases y, la segunda, donde se realizaron los reportes y protocolos verbales, se llevó a cabo en seis fases (ver Tabla 7). Cada una de las dos etapas del proyecto se enfocó en conseguir distintos tipos de evidencias de validez.

---

<sup>4</sup> De ahora en adelante se les abreviará como Excale 09

<sup>5</sup> En un principio se consideró también averiguar la presencia de sesgo por lengua materna, modalidad educativa y ruralidad; sin embargo, se descartaron por la inviabilidad para incluirlas en las dos etapas de este proyecto, principalmente por la falta de recursos humanos y económicos para concluir el estudio en los tiempos de una investigación doctoral.

**Tabla 7. Modelo metodológico del análisis de sesgo**

Etapa	Fase
1. Funcionamiento diferencial de ítems (DIF)	1. Revisión de propiedades psicométricas de los ítems 2. Cálculo del Nivel Socioeconómico 3. Análisis DIF por sexo mediante RASCH-DIF 4. Análisis DIF por Nivel Socioeconómico
2. Reportes y protocolos verbales	1. Diseño 2. Pilotaje 3. Aplicación de reportes y protocolos verbales 4. Procesamiento de la información 5. Análisis cognitivo de cada ítem 6. Análisis de sesgo mediante el Diseño Universal de las Evaluaciones

Fuente: elaboración propia

### 3.1. Etapa 1. Funcionamiento Diferencial de Ítems

La primera etapa se centró en identificar si los ítems del Excale 09 presentan un funcionamiento diferencial por sexo y/o NSE; este tipo de información corresponde a la cuarta fuente de evidencias de validez llamada relaciones con otras variables y se realizó desde un enfoque de equidad en las evaluaciones (AERA-APA-NCME, 2014). El diseño de la investigación de la primera etapa del estudio es un análisis secundario de datos; este tipo de diseños recuperan la información ya existente, generalmente en bases de datos, para realizar análisis con un fin específico que suele ser distinto al del estudio original (Cea D' Ancona, 2001).

En 2012, cuando se aplicó por última vez el Excale 09 había en todo el país 36,563 escuelas secundarias que funcionaban bajo cinco modalidades: generales, técnicas, telesecundarias, comunitarias y para trabajadores (Robles, Hernández, Zendejas y Pérez, 2013, p. 66). En la primera etapa se trabajó con las bases de datos que el Instituto Nacional para la Evaluación de la Educación (INEE) pone a

disposición del público interesado en su portal web. Dichas bases contienen la información de los resultados de los 94,269 estudiantes a los que se les aplicó el Excale 09 en 2012, los cuales pertenecían a 3,590 secundarias de todo el país. Para esta investigación, se utilizaron dos bases de datos:

- La que contiene la información del *Cuestionario para el Alumno* que se aplica a la totalidad de alumnos que respondieron el Excale 09 y que recolecta información sobre el contexto social, familiar y escolar en que viven los estudiantes.
- La que incluye los resultados de los 27,009 estudiantes que respondieron el Excale 09 de Matemáticas. Para cada alumno se informa si Acertó o Falló a determinado ítem.

La primera etapa de esta investigación dedicada a la detección de ítems con un funcionamiento diferencial, está dividida en cuatro fases concatenadas: revisión de propiedades psicométricas de los ítems, cálculo del NSE, análisis DIF y NUDIF por sexo y análisis DIF y NUDIF por NSE.

### **3.1.1. Fase 1. Revisión de propiedades psicométricas de los ítems**

En el INEE, como parte del proceso de elaboración de los Excale, se llevan a cabo análisis psicométricos de los ítems tanto en su versión para pilotaje como en su versión final. En esta fase, se aprovecharon los estadígrafos que el INEE ha generado sobre cada uno de los ítems desde la Teoría de Respuesta al Ítem

(TRI), tales como los valores infit y oufit de análisis Rasch; todos ellos, desglosados por sexo.

Esta primera fase es un prerequisite para realizar análisis posteriores. Su utilidad radica en que a partir de la información psicométrica que el INEE proporciona, se hizo un primer acercamiento (incipiente aún) sobre el sesgo de algunos ítems hacia determinadas subpoblaciones de evaluados.

### **3.1.2. Fase 2. Cálculo del nivel socioeconómico**

El NSE es uno de los dos ejes de los análisis DIF realizados en este trabajo. Como en la información pública que proporciona el INEE a los investigadores no incluye el índice socioeconómico, en esta investigación se realizó su cálculo a partir de la base de datos que contiene la información del *Cuestionario para el Alumno*<sup>6</sup>, seleccionando aquellas variables que versan sobre el contexto social y escolar en que viven los estudiantes (véase Tabla 8), y además, que dichas variables fueran las mismas que el INEE ha utilizado para construir índices sobre el NSE o capital cultural en otros estudios (Cfr. Sánchez y Andrade, 2009), con la finalidad de que el procedimiento fuera comparable. Se dividió a los estudiantes en tres estratos (bajo, medio y alto) a partir de un índice que se construyó mediante el modelamiento de crédito parcial del análisis Rasch utilizando el software Winsteps versión 3.91.

---

<sup>6</sup> Instrumento de 39 ítems que recolecta información sobre el contexto social, familiar y escolar en que viven todos los estudiantes que respondieron cualquiera de las versiones o formas del Excale 09. Los ítems fueron de selección de respuesta.

En primera instancia se puso a prueba la unidimensionalidad del constructo mediante la verificación de los índices de ajuste interno (Infit) y externo (Outfit) con valores mayores a 0.8 y menores a 1.3 lógitos. Derivado de esa primera revisión, se identificó que la variable “Hacinamiento”, construida a partir del cociente del ítem AS001 / AS002, no era parte de la variable latente NSE. Una vez eliminada esa variable, los índices de bondad de ajuste fueron los que se presentan en las dos columnas a la derecha de la Tabla 8.

Calibradas las variables, se procedió a calcular el índice socioeconómico a partir del nivel de habilidad del modelo Rasch. El índice se estandarizó con media de cero y desviación estándar de 1 logito. Se crearon tres NSE, cuyos intervalos son de igual amplitud, a partir del rango de la escala que osciló de -5.23 a 4.73 lógitos.



**Tabla 8. Variables empleadas e índices de bondad de ajuste para construir el nivel socioeconómico**

Código Ítem	Ítem	Categorías de respuesta	INFIT MNSQ	OUTFIT MNSQ
Hacinamiento	AS001 Includyéndote, ¿cuántas personas viven en tu casa? / AS002 En tu casa, ¿cuántos cuartos se utilizan para dormir?	1=1 por habitación; 2=2 por habitación; 3=3 por habitación; 4=4 o más por habitación	1.76	1.93
AS003_RE	Aproximadamente, ¿cuántos libros hay en tu casa? (No incluyas revistas, periódicos, ni tus libros escolares)	0=Ninguno; 1=1-5; 2=6-10; 3=11-15; 4=16-20; 5=21-40; 6=41-60; 7=61-80; 8=81-100; 9=101 o más	1.21	1.28
AS008_RE	¿Cuántas veces saliste con tu familia a vacacionar en los últimos dos años?	0=Ninguna; 1=1 o 2 veces; 2=3 o 4 veces; 3=5 o 6 veces; 4=7 veces o más	1.28	1.32
AS009_RE	¿Cuántos automóviles tienen en tu casa?	0=Ninguno; 1=1; 2=2; 3=3; 4=4 o más	1.17	1.27
AS011	¿Hay teléfono en tu casa?	0=No; 1=Sí	0.97	0.94
AS012	¿Hay computadora en tu casa?	0=No; 1=Sí	0.81	0.76
AS014_RE	¿Cuál es el nivel máximo de estudios que concluyó tu mamá?	0=No fue a la escuela; 1=Primaria; 2=Secundaria; 3=Bachillerato o Preparatoria; 4=Universidad	0.84	0.84
AS015_RE	¿Cuál es el nivel máximo de estudios que concluyó tu papá?	0=No fue a la escuela; 1=Primaria; 2=Secundaria; 3=Bachillerato o Preparatoria; 4=Universidad	0.83	0.82
AS016	¿Hasta qué nivel educativo te gustaría estudiar?	0=Secundaria; 1=Bachillerato o Preparatoria; 2=Carrera técnica; 3=Licenciatura o carrera universitaria; 4=Posgrado (maestría o doctorado)	1.02	1.02
AS017	¿Hasta qué nivel educativo les gustaría a tus padres que estudiaras?	0=No sé; 1=Secundaria; 2=Bachillerato o Preparatoria; 3=Carrera técnica; 4=Licenciatura o carrera universitaria; 5=Posgrado (maestría o doctorado)	0.92	0.92

**Fuente:** elaboración propia a partir de las bases de datos públicas del INEE.

### 3.1.3. Fase 3. Aplicación del modelo RASCH-DIF por subpoblación según sexo

La tercera fase consistió en analizar la presencia de DIF y NUDIF entre hombres (grupo de referencia) y mujeres (grupo focal) mediante la técnica RASCH-DIF. Como ya señaló en el segundo capítulo de esta tesis, esta técnica resulta muy apropiada para detectar DIF y NUDIF en ítems que forman parte de una prueba matricial, porque al ser parte de las técnicas que se incluyen dentro de la TRI, permiten calcular el nivel de dificultad de cada reactivo de forma independiente del nivel de habilidad de los sustentantes (Hidalgo y Gómez-Benito, 2010). Además, esta técnica es más robusta que las técnicas tradicionales para la detección de DIF como Mantel–Haenszel, la Ji-cuadrada de Lord o la comparación de niveles de dificultad de Angoff (Fidalgo, 1997).

Para el cálculo del DIF se utilizó el software Winsteps versión 3.91. En primera instancia se calcularon bajo el modelo Rasch la dificultad de cada ítem ( $\delta_r$ ) y la habilidad de cada sustentante ( $\beta_s$ ) mediante la fórmula:

$$P_i(x=1,2,\dots,n|\beta_s) = e^{(\beta_s - \delta_r)} / (1 + e^{(\beta_s - \delta_r)}); \quad (5)$$

donde  $P_i(x=1|\beta_s)$  es la probabilidad de que un sustentante seleccionado aleatoriamente responda correctamente al ítem  $i$ ;  $n$  es el número de ítems en el examen; y  $e$  la constante 2.71828. En una segunda instancia se calcularon los contrastes del modelo RASCH-DIF a partir de la fórmula:

$$\text{DIF-Contrast} = \delta_{iH} - \delta_{iM}; \quad (6)$$

donde  $\delta_{iH}$  es el nivel de dificultad del ítem  $i$  para los hombres y  $\delta_{iM}$  es el nivel de dificultad del mismo ítem  $i$  para mujeres.

Una de las recomendaciones en el análisis DIF es que se utilicen al menos dos técnicas para su detección, y a partir de la concurrencia de los resultados se diagnostique aquellos reactivos que en varios métodos son marcados (Camilli y Shepard, 1994). En atención a esa recomendación, en esta investigación se utilizó de forma adicional al método RASCH-DIF calculado mediante el software Winsteps, la metodología del RASCH-DIF mediante el método de Estimación Marginal de Máxima Verosimilitud (EMMV) utilizando la librería TAM, de reciente creación, en el software R Statistical versión 3.3.1 (Kiefer, Robitzsch y Wu, 2016).

El procedimiento para la detección del DIF, se llevó a cabo mediante la siguiente secuencia (véase el script en el Anexo 2):

1. Calibración separada, donde se estiman los parámetros de dificultad de cada uno de los reactivos  $\hat{\delta}_{i,R}$  y  $\hat{\delta}_{i,F}$  para cada una de las subpoblaciones (de referencia y focal) junto con sus respectivos errores estándar  $\sqrt{\hat{V}(\hat{\delta}_{i,R})}$  y  $\sqrt{\hat{V}(\hat{\delta}_{i,F})}$  restringiendo, además, los promedios de dificultad para que sumen cero y tengan la misma escala, es decir:  $\sum_{i=1}^N \hat{\delta}_{i,R} = \sum_{i=1}^N \hat{\delta}_{i,F} = 0$
2. Estimación del tamaño de las diferencias: Se calcula la diferencia de los parámetros de dificultad para cada uno de los reactivos a través de las subpoblaciones, es decir:  $\hat{\Delta}_i = \hat{\delta}_{i,R} - \hat{\delta}_{i,F}$

3. Índice Rasch-DIF: Se calcula el índice Rasch-DIF para cada uno de los

reactivos, a partir de: 
$$Z_i = \frac{\hat{\Delta}_i}{\sqrt{\hat{V}(\hat{\delta}_{i,R}) + \hat{V}(\hat{\delta}_{i,F})}}$$

#### **3.1.4. Fase 4. Aplicación del modelo RASCH-DIF por subpoblación según NSE**

La tercera fase consistió en analizar la presencia de DIF y NUDIF entre estudiantes de NSE Bajo (grupo focal) y NSE Alto (grupo de referencia) utilizando el mismo software y procedimiento que en la segunda fase. En los dos ejes de análisis ya señalados (sexo y NSE) los análisis DIF y NUDIF se realizaron para cada una de las tres subescalas que mide el Excale 09 de Matemáticas: sentido numérico y pensamiento algebraico (SNPA), forma, espacio y medida (FEM) y manejo de la información (MI).

Para marcar los ítems como sospechosos de sesgo por presentar DIF o NUDIF se utilizó el siguiente criterio: debieron presentar DIF-Contrast o NUDIF  $\geq 0.64$  lógitos. Se precisa que, en el caso de NUDIF, se requiere que en uno de los terciles en que se subdividió la población según su habilidad matemática se presente DIF-Contrast severo a favor del grupo focal y en otro de los terciles se registre DIF-Contrast favoreciendo al grupo de referencia, o viceversa.

### **Recursos necesarios**

Para llevar a cabo la primera etapa del estudio, se requirió la tabla de contenidos de los Excale 09 de Matemáticas, así como las bases de datos con los resultados de logro y los cuestionarios de contexto. Además de la paquetería Winsteps y librería TAM del software R.

### **3.2. Etapa 2. Análisis de sesgo mediante reportes y protocolos verbales**

En la segunda etapa del estudio se analizaron en profundidad los ítems de Matemáticas que fueron seleccionados a partir de la etapa anterior. Los ítems elegidos son aquéllos donde existe sesgo potencial para algunos de los subgrupos de análisis; esta información abonó a la obtención de evidencias de validez sobre los procesos de respuesta.

En esta etapa se utilizaron los métodos de reportes y protocolos verbales para analizar el proceso de respuesta que emplean estudiantes de tercero de secundaria a partir de la aplicación técnicas de pensamiento en voz alta y entrevistas concurrentes. La segunda etapa del estudio se desarrolló a partir de seis fases: 1) diseño, 2) pilotaje, 3) realización de reportes y protocolos verbales, 4) procesamiento de la información, 5) el análisis cognitivo de cada ítem y, 6) análisis cognitivo por cada dimensión del DUE. A continuación se describe cada una de ellas.

### 3.2.1. Fase 1. Diseño

En la primera fase, en la que se diseñó el estudio cognitivo, se desarrollaron las siguientes once actividades:

#### *A) Selección de ítems del estudio cognitivo*

Se seleccionaron 18 ítems del Excale 09 de Matemáticas. Como se muestra en el capítulo de resultados, estos ítems son los que presentaron un funcionamiento diferencial severo entre las subpoblaciones utilizadas como eje de análisis. Estos ítems y las especificaciones a partir de las cuales se elaboraron, fueron facilitados por el INEE a reserva de confidencialidad, por lo cual varios aspectos de dichos elementos no se presentan en la tesis.

#### *B) Identificación de los procesos cognitivos necesarios para responder a los ítems, declarados en el modelo teórico de la prueba*

Para realizar esta actividad, se analizaron las especificaciones de cada uno de los ítems que presentaron DIF severo en la primera etapa del estudio. Las especificaciones de los reactivos es un término del argot de los diseñadores de pruebas que refiere a las indicaciones pormenorizadas a partir de las cuales se diseñará un grupo de ítems en específico y los cuales pretenden medir un contenido y/o habilidad en particular. Las especificaciones de los Excale 09 están conformadas por seis apartados; dentro del segundo apartado titulado *Descripción del contenido curricular a evaluar* existen dos subapartados que resultaron fundamentales para realizar esta actividad: los nombrados *Actividades*

*cognoscitivas involucradas en la contestación correcta del ítem y Conocimientos y habilidades previos requeridos para contestar el ítem correctamente.*

En el subapartado *Actividades cognoscitivas involucradas en la contestación correcta del ítem* se establecen el o los procesos cognitivos que los sustentantes tendrían que usar para responder correctamente el ítem en cuestión. Por ejemplo, en una de las especificaciones para resolver problemas aditivos con números fraccionarios con distinto denominador se señala que las actividades cognoscitivas que los estudiantes tendrán que poner en juego son:

- a) verificar que el problema presentado se resuelve mediante la suma o resta de números fraccionarios con distinto denominador; b) aplicar el algoritmo de la equivalencia de fracciones en la adición o sustracción de números fraccionarios con distinto denominador, y c) resolver problemas por medio de una(s) operación(es) que implique(n) el uso de números fraccionarios con distinto denominador, a partir de una situación dada (INEE, 2010a, s/p).

En el subapartado *Conocimientos y habilidades previos requeridos para contestar el ítem correctamente*, se establecen los saberes matemáticos que son condición necesaria para que los alumnos respondan de manera correcta el ítem. Por ejemplo, para la misma especificación nombrada *Resolver problemas aditivos con números fraccionarios con distinto denominador* se indica que los conocimientos y habilidades previas son:

- a) otorgar sentido a cantidades expresadas como fracción; b) interpretar cantidades fraccionarias en contextos de medición; c) expresar cantidades utilizando fracciones; d) emplear los algoritmos de la adición, sustracción de números naturales mediante

equivalencias de fracciones; e) resolver, con base en la equivalencia de fracciones, problemas aditivos simples y combinados, directos e inversos que involucren números fraccionarios; y f) identificar y producir familias de fracciones equivalentes mediante la amplificación y simplificación, reconociendo que todas ellas expresan una misma cantidad (INEE, 2010a, s/p).

Así, para cada ítem se identificaron los procesos cognitivos que desde el diseño de la prueba debían utilizar los sustentantes.

*C) Anticipación de los procesos cognitivos necesarios para responder a los ítems durante las entrevistas*

Tomando como referente las especificaciones de ítems, un par de especialistas en enseñanza y evaluación de las matemáticas en educación básica, definió por cada ítem el proceso cognitivo principal y, de ser el caso, sus variantes por las cuales deberían transitar los alumnos para responderlo de forma correcta. Asimismo, las especialistas ayudaron a identificar cuáles son las distorsiones más comunes asociadas a determinadas respuestas incorrectas. Como resultado de este análisis detallado se obtuvo el o los procesos cognitivos (desde un punto de vista teórico) que utilizarían los entrevistados para responder cada ítem.

Para realizar su trabajo, se tuvo una reunión inicial de trabajo donde se les presentaron los objetivos del proyecto de investigación, sus etapas y sus fases, de tal manera que tuvieran una visión global del estudio, del papel que ellas desempeñarían y de cómo se utilizaría la información que se generara producto de su trabajo. Una vez que aceptaron participar en la investigación, se les pidió que



firmaran una carta-compromiso de confidencialidad, pues los ítems y las especificaciones que se les proporcionaron aún no están liberados por el INEE y forman parte de su banco de ítems vigente.

Para guiar el trabajo de las especialistas, se diseñó un formato (ver Anexo 3) en el que registraron: a) los datos de identificación del ítem a analizar, b) los conocimientos y habilidades previos requeridos para contestar el ítem correctamente, c) las actividades cognoscitivas involucradas en la contestación correcta del ítem, d) el proceso cognitivo que anticiparon que los entrevistados emplearían para enfrentarse a la tarea evaluativa, para lo cual tuvieron que llenar un diagrama que representase el sendero cognitivo previsto (lineal, en paralelo, en espiral, lineal en un inicio y luego una bifurcación, etc.), y e) el nivel cognitivo que demanda el ítem (conceptual, procedimental, resolución de problemas conocidos o típicos, y resolución de problemas desconocidos o transferencia), para lo cual se les proporcionó la definición de cada nivel según el marco de referencia del Excale (ver Anexo 4).

Respecto al perfil de las especialistas, una de ellas es egresada de un doctorado en enseñanza de las matemáticas por el Departamento de Matemática Educativa del Cinvestav, además de realizar proyectos de investigación sobre matemática educativa, colabora en la edición de la Revista Educación Matemática, lo que prueba su *expertise* en trabajos de investigación de este campo. La segunda especialista en enseñanza de las matemáticas es egresada de una normal superior donde se especializó en enseñanza de las matemáticas, obtuvo una maestría en investigación educativa y cursa un doctorado en investigación

educativa donde realiza un proyecto de investigación sobre observación de la práctica docente en la asignatura de matemáticas. Su adecuación para colaborar como especialista se sustenta en su trabajo en varios proyectos de investigación con docentes de matemáticas, de evaluación de los aprendizajes y su formación de pregrado.

Considerando el importante trabajo intelectual que implicó anticipar el sendero cognitivo de cada ítem, se consiguieron recursos para ofrecerles una compensación, lo que a su vez incrementó las probabilidades de que se realizara este proceso con la rigurosidad que ameritó. En el E-anexo 1 se presentan el producto de la especialista 1 y en el E-anexo 2 el trabajo de la segunda especialista.

*D) Selección de la modalidad de técnica cognitiva a emplear*

Existe una amplia variedad de modalidades bajo las cuales se han realizado los protocolos verbales, sin embargo, como señala Willis (2015) pueden englobarse en dos grandes tipos: aquéllas que aluden al pensamiento en voz alta o *think aloud* (donde se le pide a los participantes que vayan expresando todo lo que piensan mientras responden los ítems) y las que utilizan el sondeo verbal (*verbal probes*) mediante el cual se pide a los evaluados que aporten información adicional acerca de sus pensamientos al enfrentarse a un ítem o un conjunto de ellos.

A diferencia de la modalidad *think aloud* que necesariamente ocurre al mismo tiempo que se resuelve un ítem, el sondeo verbal puede ser concurrente o

retrospectivo. Es concurrente cuando poco tiempo después (inclusive inmediatamente después) de que el examinado responde el ítem en cuestión, se le pide que explique, describa, profundice o haga explícito el proceso mental que siguió para responderlo. En cambio, el sondeo verbal retrospectivo es una indagación que se lleva a cabo después de un tiempo considerable de que el sustentante se enfrentó a las tareas evaluativas.

En este estudio se emplearon la modalidad de pensamiento en voz alta y el sondeo verbal concurrente. Aunque se privilegió identificar el modelo cognitivo empleado por los sustentantes a través de la verbalización de sus pensamientos mientras se enfrentan a cada ítem, resultó conveniente utilizar como estrategia complementaria el sondeo verbal, en especial para identificar comprensiones de conceptos implicados en los ítems y que estén asociados a aspectos culturales y de su NSE, que es el foco de este trabajo.

*E) Diseño de la Guía para la realización de entrevistas concurrentes y la técnica de pensamiento en voz alta*

La Guía es la que pautó la intervención del entrevistador durante el estudio cognitivo (ver Anexo 5). En ella se detallan los diferentes pasos que debieron seguirse antes, durante y después de realizar la técnica de pensamiento en voz alta y las entrevistas concurrentes con cada alumno; de forma especial, se presentan las indicaciones que se les dieron a los entrevistados a fin de entrenarlos para pensar en voz alta, se incluyeron las consignas que se les dieron a los participantes cuando dejaban de verbalizar lo que estaban pensando mientras resolvían los ítems, se definieron las posibles preguntas que se utilizaron

en las entrevistas concurrentes (tanto para perfeccionar la ruta cognitiva que utilizó el estudiante como para identificar fuentes de sesgo asociadas al NSE del que proviene el participante) así como las especificidades que el entrevistador debió cuidar con ciertos ítems.

*F) Diseño del instrumento para recolectar los datos*

Si bien las entrevistas concurrentes y la técnica de pensamiento en voz alta se videograbaron de tal manera que hubiera un registro de la información obtenida, se planeó que también hubiera un instrumento que permitiese recolectar cierta información, básica aún, pero de primera mano mientras se realizaba el levantamiento de datos. Para ello se diseñó el *Formato de Recolección de Datos* (ver Anexo 6) que el entrevistador completó para cada uno de los ítems. En dicho formato se registraron: palabras leídas de manera incorrecta por el entrevistado, términos en los que el estudiante manifestó confusión o desconocimiento, preguntas de sondeo concurrente adicionales a las previstas en el protocolo y que fueron hechas por el entrevistador, si el entrevistado respondió correctamente el ítem, si el proceso para enfrentarse a la tarea evaluativa fue apropiado, entre otras cuestiones. Este instrumento sirvió como un primer análisis de cada ítem, que se complementó con la revisión de las videograbaciones.

*G) Definición de los instrumentos para el procesamiento de la información*

Para procesar la información obtenida a través de la videograbación y del *Formato de Recolección de Datos* que completó el entrevistador, se diseñó el *Formato para concentrar y procesar la información de cada ítem-estudiante* (ver

E-anexo 3), que se detalla en el apartado donde se describe la fase del procesamiento de la información.

#### H) *Selección de los participantes*

La selección de los estudiantes se hizo a partir de los siguientes criterios: el primero y fundamental fue su pertenencia a una de las tres subpoblaciones participantes de NSE (alto, medio o bajo) en las que se presumió sesgo en los ítems como resultado de los análisis DIF llevados a cabo en el primer estudio. Un segundo criterio de selección fue que los participantes provinieran de distintas modalidades en la que se ofrece la educación secundaria en México, lo que ayudó a aleatorizar las variables no consideradas como criterio de selección. El tercer criterio fue que en ninguna ocasión anterior hayan respondido los Excale 09. El cuarto criterio fue su rendimiento académico en Matemáticas representado por sus calificaciones escolares en esta asignatura, así, a sus docentes se les pidió que escogieran un alumno con promedio de 9 o superior y otro alumno con promedio de 7 o menos, de tal forma que cuando ellos se enfrentaran a las tareas evaluativas, hubiera mayor varianza de los diferentes niveles de logro, y se pudiera explorar la presencia de sesgo parcial o específico para ciertos grupos.

Se entrevistaron seis estudiantes de tercero de secundaria y sus tres profesores de matemáticas. Se seleccionó un par de alumnos de una Telesecundaria de NSE bajo y el profesor que les impartía clases; asimismo, se seleccionaron dos estudiantes de una secundaria general pública de NSE medio y a su maestra de matemáticas; y, finalmente, se seleccionaron dos alumnos de una secundaria privada de NSE alto y a su profesor de matemáticas (ver Tabla 9).

**Tabla 9. Cantidad de participantes en el estudio cognitivo, según su nivel socioeconómico y la modalidad de secundaria en la que estudian**

Nivel Socioeconómico	Modalidad						Total
	Telesecundaria		General-Pública		General- Privada		
	Alumno	Profesor	Alumno	Profesor	Alumno	Profesor	
Alto	0	0	0	0	2	1	3
Medio	0	0	2	1	0	0	3
Bajo	2	1	0	0	0	0	3
Total	3		3		3		9

Fuente: elaboración propia

*I) Definición de la cantidad de ítems por participante*

Considerando que dentro de cada grupo de NSE hubo sólo un estudiante de rendimiento alto y un estudiante de rendimiento bajo, se definió que cada entrevistado respondiera los 18 ítems que mostraron DIF severo por NSE, para que ante una eventual situación donde los resultados requirieran explorar si el factor asociado a los resultados diferenciales fuera el rendimiento escolar, se tuviera la posibilidad de hacer dicha comparación. Así, al final se tuvieron 108 respuestas de alumnos y 54 respuestas de los profesores, siendo un total de 162 respuestas.

*J) Preparación de cuadernillos para estudio cognitivo*

Los 18 ítems que se aplicaron en el estudio cognitivo son sólo una muestra de la escala completa de Matemáticas de los Excale 09 que consta de cien ítems. Por lo cual se requirió preparar los cuadernillos de la prueba para que contuvieran únicamente los 18 ítems a probar. Además, como se pretendió que no hubiera distractores que alteraran los resultados de los reportes y protocolos verbales, se

cuidó que en cada hoja hubiera un solo ítem y que el cuadernillo estuviera armado de tal forma que el participante sólo pudiera trabajar en un ítem a la vez.

#### *K) Adaptación del laboratorio cognitivo*

Se adaptó provisionalmente un cubículo de la UAA (en el edificio 13) como laboratorio cognitivo; se dispuso de mobiliario (escritorio y sillas) para que cada entrevistado respondiera los ítems, se dispuso de una videocámara de alta definición que fue colocada en un trípode y tenía conectado el receptor de un micrófono lavalier inalámbrico cuyo transmisor fue colocado a cada estudiante, de tal manera que tanto sus verbalizaciones mientras respondía los ítems, así como sus respuestas a las entrevistas concurrentes, fueran grabadas con toda claridad.

### **3.2.2. Fase 2. Pilotaje**

La estrategia señalada en la fase del diseño se puso a prueba con un estudiante de tercero de secundaria. El propósito central de esta fase consistió en identificar y mejorar el proceso de entrenamiento que recibieron los participantes para responder a las entrevistas concurrentes y la técnica de pensamiento en voz alta, evaluar las técnicas mismas e instrumentos de obtención de información así como la operación del laboratorio cognitivo. A partir del piloto, se hicieron ajustes en la *Guía para la realización de entrevistas mediante protocolos verbales y técnicas de pensamiento en voz alta* así como en el *Formato para recolección de información* precisando algunas de las instrucciones, y anticipando la necesidad de algunos

materiales como borrador, sacapuntas, agua embotellada y un paquete de materiales como gratificación para los docentes.

### **3.2.3. Fase 3. Realización de reportes y protocolos verbales**

La tercera fase consistió en llevar a cabo los reportes y protocolos verbales, los cuales estuvieron pautados por la *Guía para la realización de entrevistas mediante protocolos verbales y técnicas de pensamiento en voz alta*. Durante esta fase se siguieron los siguientes pasos:

1) Revisar instalación del laboratorio cognitivo:

- Mobiliario: 1 mesa de trabajo o escritorio y 2 sillas; ubicadas de frente una con otra y a no más de 1.5 mts. de distancia entre ellas.
- Equipo de videograbación: videocámara, trípode, micrófono lavalier, memoria SD, cargador eléctrico de la videocámara, extensión eléctrica.
- Preparación del equipo de videograbación en su ubicación apropiada (evitando contraluz y que aparecieran en la toma tanto el entrevistado como el entrevistador), así como realizar todas las conexiones entre los diferentes componentes de la videograbación, de tal manera que estuviera lista para su utilización.

2) Revisión de materiales necesarios para realizar la entrevista:

- Material para entrevistado: cuadernillo con ítems, lápiz, borrador, sacapuntas, consentimiento informado, hojas blancas.
- Material para entrevistador: guía de realización de entrevistas, formato de recolección de datos, ítems para realizar entrenamiento, lápiz y sacapuntas.



3) Bienvenida al entrevistado

- Breve presentación personal por parte del entrevistador.
- Agradecimiento al entrevistado por participar en el estudio.
- Breve presentación del estudio (sobre todo con aquellos participantes que no estuvieron durante la negociación de su participación).
- Descripción al entrevistado de actividades que llevará a cabo durante la sesión.

4) Firma de consentimiento informado (opcional, cuando esto no haya ocurrido durante la negociación de su participación).

En concordancia con los criterios éticos de esta investigación, a los participantes se les solicitó su consentimiento para su participación en el estudio. Como parte de esta actividad, se les proporcionó información sobre los propósitos del estudio, la manera en que se utilizarían los resultados, el papel que ellos jugaron dentro de la investigación, así como los cuidados que se tendrían para salvaguardar la confidencialidad de la información que proporcionen. Para ello, se elaboró una carta de consentimiento (ver Anexo 7) que los participantes debieron firmar como condición para participar. En algunos casos el consentimiento informado ocurrió durante la fase en la que se invitó a los entrevistados a colaborar en el estudio.

5) Entrenamiento con las técnicas a emplear

- Al entrevistado se le leyó el siguiente mensaje:

*“Estamos interesados en conocer cómo respondes algunas preguntas de opción múltiple de un examen de matemáticas. Para ello, voy a mostrarte*

*unos ítems y te pediré que pienses en voz alta cómo los vas solucionando, después te podría hacer algunas preguntas. Pensar en voz alta significa que quiero que me digas todo lo que pasa por tu mente desde el primer momento en que veas la pregunta del examen hasta el final cuando elijas la opción que consideres correcta. Mientras revisas cada pregunta, es muy importante que te mantengas hablando, si en algún momento guardas silencio, te pediré que vuelvas a pensar en voz alta. Actúa como si estuvieras solo(a) en este salón y hablando para ti mismo(a). ¿Hasta ahora todo lo que te comenté es claro?”*

- Al entrevistado se le ofreció un ejemplo de cómo pensar en voz alta. Para ello, el entrevistador ejemplificó cómo resuelve alguna operación (resta o multiplicación) de dos dígitos. Para ello, fue importante que durante todo el proceso de respuesta, el entrevistador verbalizara lo que pensaba.
- Se realizó la primera práctica de pensamiento en voz alta. Resolviendo operaciones (no ítems todavía). Al entrevistado se le leyó:

*Ahora vamos a practicar un poco, para que te acostumbres a pensar en voz alta. Te voy a pedir que esta hoja blanca que te acabo de proporcionar multipliques dos números. Así que piensa en voz alta mientras multiplicas 15 por 25.*

Por su parte, el entrevistado propició que el estudiante verbalizara el proceso utilizado.

- Al entrevistado se le dio un ejemplo de cómo resolver un ítem. Para ello el entrevistador respondió uno o dos reactivos a modo de ejemplo frente al

participante para que pudiera observar el comportamiento y desempeño que se esperaba de él. Para esta etapa de entrenamiento, se utilizaron algunos de los ítems del Excale 09 que no tuvieron DIF, para no distorsionar los procesos cognitivos que el estudiante empleaba en cada uno de los ítems que interesó analizar. El entrevistador leyó las siguientes instrucciones:

*“Ahora voy a pensar en voz alta mientras resuelvo este problema. Eso significa que voy a decir todo lo que pasa por mi mente, incluyendo cómo entendí el problema que tengo que resolver, si hay palabras que no comprendo, si me resultan claras las indicaciones, si el problema presenta los datos suficientes para resolverlo, si recuerdo algún procedimiento que pueda utilizar para resolverlo, si en las opciones de respuesta viene alguna palabra que no comprendo o tengo dudas de lo que significa, etc.”*

El entrenamiento terminó cuando el investigador observó que el participante presentó un nivel suficiente de desautomatización<sup>7</sup> y de conciencia de los procesos y recursos cognitivos que usaba para responder un ítem; de tal forma que en las verbalizaciones del participante se pudieran identificar de forma clara los pasos que utilizó para responder el ítem y, de ser el caso, las dificultades que tuvo para su resolución.

- Cierre del entrenamiento. Una vez concluido el entrenamiento, se le recordó al entrevistado que el principal objetivo por el que se le estaba

---

<sup>7</sup> La desautomatización de las respuestas consiste en que el entrevistado se vuelva consciente del proceso cognitivo que emplea para resolver un problema y, posteriormente, sea capaz de expresar los pasos que siguió para responder a un ítem.

entrevistando es conocer el proceso que empleó para responder cada ítem.

Al final, se le preguntó si tenía alguna duda, eventualmente se le resolvió y, ante la ausencia de dudas, se continuó con el siguiente paso.

6) Aplicación de técnica de pensamiento en voz alta y entrevista concurrente.

- Se inició la videograbación, cerciorándose que el micrófono lavalier estuviera funcionando.
- Se entregó el cuadernillo al estudiante y se le pidió que leyera las indicaciones generales, donde se le hizo saber que en cada página sólo encontrará un ítem, y que no podría cambiar de página hasta que el investigador se lo haga saber.
- Para cada ítem se realizó el siguiente ciclo. Es decir, se repitió 18 veces:
  - a. Se le pidió al estudiante o profesor entrevistado que verbalizara sus pensamientos. Durante este tiempo el entrevistador tuvo principalmente dos funciones: i) asegurarse que el entrevistado piense en voz alta mientras responde el ítem e, ii) identificar lagunas de información para registrar en el *Formato de recolección de datos* las preguntas adicionales que formularía posteriormente para subsanar dichos vacíos de información.
  - b. Una vez que el estudiante registró en el cuadernillo su respuesta a un ítem en específico, se realizó la entrevista concurrente a partir de las preguntas eje incluidas en el *Formato de recolección de datos* y de preguntas adicionales que el investigador fue redactando a partir de la técnica de pensamiento en voz alta; todo ello con el objetivo de

detallar el proceso empleado y, en especial, identificar ciertos términos que por la forma en que se comprenden, pudieran alterar el resultado del ítem.

- 7) Agradecimiento al entrevistado. Donde además de un agradecimiento verbal, a los estudiantes se le gratificó con una batería portátil, a modo de obsequio; a los docentes, de manera adicional se les regaló un paquete de materiales con temáticas para el mejoramiento de las prácticas de evaluación el aula.
- 8) Integración del expediente de cada entrevistado. El cual debería contener:
  - *Formato de recolección de datos* debidamente complementado.
  - Archivo con la videograbación. El cual se nombró con el mismo número de folio del entrevistado.
  - El cuadernillo del examen respondido por el entrevistado.
  - Consentimiento informado, firmado por el entrevistado.

#### **3.2.4. Fase 4. Procesamiento de la información**

Para procesar la información obtenida a través de las entrevistas, se utilizó el *Formato para concentrar y procesar la información de cada ítem-estudiante* (ver E-anexo 3). La información de dicho formato varía según el ítem a analizar, por lo cual se diseñaron 18 formatos y cada uno está compuesto por doce apartados:

1. Datos de identificación del sustentante, tales como un folio interno, subpoblación a la que pertenece, tipo de escuela de la que proviene (general pública, general privada o telesecundaria)

2. Datos del ítem, como el folio, el dominio al que pertenece y el código de la especificación de la que se creó.
3. Conocimientos y habilidades previos requeridos para contestar el ítem correctamente. Información que se retomó de la especificación del ítem.
4. Actividades cognoscitivas involucradas en la contestación correcta del ítem. Información retomada de la especificación del ítem.
5. Proceso(s) cognitivo(s) previstos por especialistas en enseñanza de las matemáticas. Que se elaboró a partir del E-anexo 1 y E-anexo 2.
6. Nivel de demanda cognitiva previsto por especialistas en enseñanza de las matemáticas
7. Proceso(s) cognitivo(s) empleados por sustentantes. Donde se registró el proceso de respuesta que empleó cada uno de los entrevistados para responder a un ítem en específico
8. ¿El entrevistado respondió correctamente el ítem? Cuyas respuestas pudieron ser “Sí” o “No”.
9. Adecuación del proceso de solución del problema.
10. Fluidez en la lectura del ítem
11. Tipo de problema que presentó el ítem. Donde se registró si el reactivo presentó alguno de los cuatro siguientes problemas y, en caso de la respuesta sea afirmativa, se registró una descripción:
  - a. Definición imprecisa del constructo
  - b. Ítem no accesible o sesgado
  - c. Procedimientos e instrucciones complejas, poco claras e intuitivas
  - d. Incomprensibilidad o problemas de legibilidad del ítem

12. Notas adicionales. En el que se pudo registrar información sobre problemas distintos a los cuatro señalados en el punto anterior, o bien, se adelantaran algunas conclusiones preliminares.

Si bien cada uno de los apartados jugó un papel en el estudio cognitivo, el apartado 11 se consideró medular, pues ahí se señaló si el entrevistado había mostrado evidencias de que el ítem en cuestión presentaba alguno de los cuatro tipos de sesgo que se exploraron en este trabajo, los cuales se definieron a partir de los siete elementos clave del Diseño Universal de las Evaluaciones (Thompson *et al.*, 2002). Como se explicó en el capítulo 2 de esta tesis, son siete los elementos clave del DUE y en esta investigación se exploraron cuatro de ellos, con una variante, en lugar de analizarlas como principios deseables, se trabajaron como su contraparte, es decir, como posibles problemas de la varianza irrelevante al constructo: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; 3) Procedimientos e instrucciones complejas, poco claras e intuitivas; y 4) Incomprensibilidad o problemas de legibilidad del ítem.

Dos de los elementos del DUE, evaluación inclusiva de la población y sensibilidad a las adaptaciones, no fueron incluidos en este trabajo porque requieren de metodologías distintas a las utilizadas en esta investigación y, además, al requerir un análisis de aspectos del diseño inicial de la prueba y de la logística de aplicación, rebasan los alcances del estudio. Por su parte, el principio del DUE que refiere a la máxima visibilidad, si bien sí es susceptible de abordarse mediante entrevistas cognitivas, no fue analizado en esta investigación por la imposibilidad para disponer de los cuadernillos originales de la prueba, que son indispensables

para evaluar este elemento, pero que por cuestiones de seguridad el INEE proporcionó sólo el fragmento del cuadernillo que contenía cada uno de los 18 ítems de matemáticas con DIF.

La evidencia se conformó a partir del pensamiento en voz alta que realizó cada entrevistado mientras respondía los ítems, así como sus respuestas a las preguntas que hizo el entrevistador a manera de sondeo concurrente para explorar algunas partes del proceso de respuesta que no fueron detallados mientras el entrevistado verbalizaba lo que pensaba para enfrentarse a la tarea evaluativa. No se hizo una transcripción completa de las entrevistas cognitivas, sino que se hizo una transcripción selectiva de las videograbaciones, de tal manera que lo transcrito ilustrara la presencia o ausencia de los cuatro tipos de sesgo que se exploraron en esta investigación.

En total, para procesar la información se complementaron 162 formatos, que provienen de los 9 entrevistados y su respuesta a cada uno de los 18 ítems. Para organizar los archivos se utilizó una nomenclatura en la que se concatenó el número del ítem y el folio del entrevistado. Por ejemplo, el archivo que contiene el procesamiento de la información del entrevistado #3 sobre el ítem SMA-05 se nombró de la siguiente manera: SMA-05\_F03.

### **3.2.5. Fase 5. Análisis cognitivo de cada ítem**

La quinta fase consistió en analizar la información procesada de los nueve entrevistados para cada uno de los 18 ítems, éstos últimos constituyeron la unidad de análisis. El análisis de cada ítem incluyó:



- La identificación del ítem. Que incluye el reactivo donde se señala cuál de las cuatro opciones es la correcta; su ubicación en el programa de matemáticas (eje, tema y subtema); la especificación del Excale a partir de la cual fue diseñado; y el tipo de DIF (uniforme o no uniforme), el sentido (positivo o negativo) y la magnitud (en lógitos) que presentó en la primera etapa de esta investigación.
- La resolución del ítem y adecuación del proceso empleado. Donde se presenta la cantidad de alumnos y docentes (desagregados por NSE) que respondieron de manera correcta e incorrecta el ítem, así como la adecuación del proceso empleado para su resolución.
- Tipo de problemas de varianza irrelevante al constructo que presenta el ítem. En el que se presenta la cantidad de entrevistados (desglosados por NSE) que aportaron evidencias para señalar que el ítem tiene alguno de los siguientes cuatro problemas de sesgo: definición imprecisa del constructo; ítem no accesible o sesgado; procedimientos e instrucciones complejas, poco claras e intuitivas; e incomprendibilidad o problemas de legibilidad del ítem.

Es importante señalar que el foco del análisis no se centró en identificar la concordancia entre el modelo teórico de la prueba y un modelo identificado a partir de los procesos de respuesta de los examinados, sino que estuvo centrado en identificar posibles fuentes de sesgo en función del NSE de los sustentantes.

### **3.2.6. Fase 6. Análisis de sesgo mediante el diseño universal de las evaluaciones**

La sexta fase consistió en la realización de un análisis de los resultados de la fase previa tomando como eje cada una de las cuatro dimensiones del DUE que se exploraron en este trabajo. Así, se realizó una síntesis para cada uno de los cuatro posibles problemas de sesgo que se exploraron: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; 3) Procedimientos e instrucciones complejas, poco claras e intuitivas; y 4) Incomprensibilidad o problemas de legibilidad del ítem.

Para la realización de la síntesis por cada tipo de problema, se contabilizó la cantidad de reactivos que lo presentaron, desglosándolo por el tipo de DIF con el que fue diagnosticado y el tipo de entrevistado. Además, se incluyeron fragmentos representativos del problema en cuestión, que fueron obtenidos a partir de los reportes y protocolos verbales.

#### **Recursos necesarios**

Para llevar a cabo la segunda etapa, se solicitó a personal del INEE los ítems de Matemáticas que fueron seleccionados en la etapa 1 de esta investigación, así como sus respectivas especificaciones. Se compró el siguiente material necesario para la implementación de esta etapa: videocámara de alta definición con micrófono inalámbrico de solapa, software especializado para el análisis de video,

computadora de alto rendimiento para manejo de video en alta definición y hardware para almacenamiento de video.

Considerando que se requirió un tiempo considerable del trabajo que realizaron las especialistas que apoyaron en la elaboración del modelo teórico de cada ítem, así como en la realización de las entrevistas cognitivas, se consiguieron recursos para darles una contribución, lo que permitió una mayor exigencia en la rigurosidad con que se desempeñaron, lo que a su vez posibilitó obtener información de mayor calidad. De la misma manera, también se obtuvieron recursos para contar el apoyo de un videograbador entrenado, con experiencia incluso como productor de televisión, lo que contribuyó a disponer de videograbación de muy alta calidad técnica. Asimismo, se consiguieron recursos para entregarles un obsequio a los entrevistados, que sirvió como incentivo durante la invitación para participar en el estudio.

#### **4. RESULTADOS DEL ANÁLISIS DIF**

En este capítulo se presentan los resultados de la primera etapa de la investigación doctoral, que consistió en el análisis de los ítems del Examen de la Calidad y el Logro Educativos (Excale) de Matemáticas para 3° de secundaria en la búsqueda de algún funcionamiento diferencial (DIF). El capítulo está organizado en cuatro apartados, en el primero se presentan sus propiedades psicométricas bajo la Teoría de Respuesta al Ítem (TRI), en el segundo se describen los resultados empíricos del análisis DIF utilizando como variable de segmentación el sexo, en el tercer apartado se muestran los resultados de los análisis DIF por nivel socioeconómico (NSE) y en el cuarto se presentan los reactivos con sospecha de sesgo.

##### **4.1. Propiedades psicométricas bajo la Teoría de Respuesta al Ítem**

Como ya se señaló en el capítulo previo, bajo la TRI dos de los estadígrafos más relevantes para valorar la bondad de ajuste de un ítem son los valores de ajuste interno (Infit) y ajuste externo (Outfit). El valor de Infit se interpreta como una media cuadrática ponderada de residuales que es sensible a patrones de respuesta irregulares. Este estadígrafo captura comportamientos de respuestas no esperadas, es decir anomalías según el modelo Rasch, a ítems calibrados cerca del nivel de habilidad del sustentante. Por su parte, el estadígrafo Outfit es el promedio de los residuales estandarizados derivados tanto de sustentantes como de ítems. Su valor se interpreta como una media cuadrática no ponderada sensible a los comportamientos extremos no esperados en los patrones de

respuesta. Este estadígrafo de ajuste es sensible a valores extremos y, aún más sensible, a comportamientos no esperados que afectan respuestas a ítems que se encuentran lejos del nivel de habilidad del sustentante (Bond y Fox, 2001, citados en González-Montesinos, 2008).

Para ambos estadígrafos, lo deseable es tener valores cercanos a uno; si el valor de Infit o Outfit es mucho mayor que 1, indica una baja discriminación del ítem, en cambio, valores por debajo del 1 son indicativos de una alta discriminación, lo que significa que los datos muestran menos variabilidad que la esperada por el modelo Rasch (OECD, 2012). Aunque no existe un solo criterio para definir una adecuada bondad de ajuste, en las pruebas del Instituto Nacional para la Evaluación de la Educación (INEE) se aceptan valores de Infit y Outfit dentro del rango de 0.8 a 1.2. (Backhoff *et al.*, 2008; OECD, 2009); no obstante, hay autores que sugieren que el intervalo puede extenderse de 0.8 hasta 1.3 (González-Montesinos, 2008).

En el caso de los Excale 09 de Matemáticas, se encontró que la gran mayoría de sus ítems están muy bien calibrados al modelo Rasch, es decir, sus valores Infit y Outfit están dentro del rango de 0.8 y 1.2, lo que indica que hay una buena discriminación de los reactivos para distinguir entre alumnos que no tienen la habilidad para responderlos correctamente y, al contrario, ser sólo respondidos por estudiantes que sí poseen la habilidad matemática para contestarlos de forma correcta.

En la subescala de *Sentido Numérico y Pensamiento Algebraico*, de los cuarenta ítems que la conforman sólo tres ítems (SMB\_03, SMC\_05 y SMG\_01) presentan valores ligeramente por encima de 1.2, pero incluso estas cifras están por debajo de 1.3 que es considerado el valor límite permitido para estos dos criterios de bondad de ajuste (ver Tabla 10).

**Tabla 10. Valores Infit y Outfit de la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas**

Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ	Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ
SMA_01	0.77	0.97	0.99	SME_03	0.06	1.07	1.1
SMA_02	-1.06	0.89	0.84	SMF_01	-0.29	1.06	1.09
SMA_05	-0.02	1.01	1.02	SMF_02	-1.46	0.92	0.87
SMA_08	-1.74	0.94	0.89	SMF_03	0.59	1.01	1.02
SMA_12	-0.1	0.95	0.93	SMF_04	-0.15	1.01	1.01
SMB_01	0.18	0.92	0.89	SMF_05	-0.11	1.01	1.03
SMB_02	0.39	0.99	0.98	SMG_01	1.15	1.13	<b>1.23</b>
SMB_03	-0.09	1.19	<b>1.26</b>	SMG_02	-0.08	0.98	0.97
SMB_04	0.88	1.11	1.17	SMG_03	0.42	1.02	1.05
SMB_05	-1.47	0.91	0.84	SMG_04	-0.83	0.96	0.94
SMC_01	-0.08	1.01	1.02	SMG_05	0.02	0.93	0.91
SMC_02	-0.4	1.05	1.07	SMG_08	0.15	0.98	0.97
SMC_03	0.43	1.02	1.02	SMG_09	-0.75	0.98	0.97
SMC_05	1.68	1.06	<b>1.25</b>	SMG_10	-0.18	0.93	0.9
SMC_06	0.02	0.93	0.9	SMH_01	-0.79	1.01	1
SMD_01	0.88	1.07	1.12	SMH_02	0.42	1.03	1.03
SMD_10	-0.65	0.94	0.92	SMH_03	0.03	1.05	1.05
SMD_12	0.72	0.93	0.92	SMH_04	1.35	1.07	1.18
SME_01	0.14	0.94	0.92	SMH_05	-0.23	0.94	0.92
SME_02	-0.52	1.15	1.19	SMH_07	0.72	0.94	0.95

**Notas:** Infit MNSQ = Ajuste interno a través de la media cuadrática no estandarizada

Outfit MNSQ = Ajuste externo a través de la media cuadrática no estandarizada

**Fuente:** elaboración propia

En la subescala de *Forma, Espacio y Medida*, de los 32 ítems que la conforman sólo tres ítems (SMD\_04, SMG\_07 y SMH\_11) presentan valores ligeramente por fuera del rango aceptable de 0.8 a 1.2 (ver Tabla 11).

**Tabla 11. Valores Infit y Outfit de la subescala *Forma, Espacio y Medida* del Excale 09 de Matemáticas**

Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ	Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ
SMA_03	0.2	0.96	0.95	SMD_09	-0.96	0.91	0.87
SMA_04	-0.09	0.9	0.88	SME_04	0.9	1.06	1.12
SMA_06	-0.04	1.04	1.06	SME_05	-0.56	1	0.99
SMA_07	-0.13	0.99	0.99	SME_06	-0.93	0.92	0.89
SMB_06	-0.11	1.01	1.01	SME_07	0.28	0.99	0.99
SMB_07	-1.28	0.97	0.95	SME_08	-0.07	1.08	1.09
SMB_08	0.61	1.05	1.06	SMF_06	0.17	1.06	1.1
SMB_09	-0.43	0.93	0.91	SMF_07	-0.37	1.08	1.11
SMB_10	0.27	1.01	1.04	SMF_08	1.27	0.97	0.98
SMD_02	1.48	0.96	0.95	SMF_10	-1.14	0.96	0.94
SMD_03	-0.97	0.96	0.95	SMG_06	-1.12	0.95	0.92
SMD_04	0.69	1.16	<b>1.26</b>	SMG_07	2.48	0.9	<b>0.77</b>
SMD_05	-0.39	1.09	1.12	SMH_08	0.67	1.07	1.11
SMD_06	0.94	0.92	0.91	SMH_09	-0.81	0.99	0.98
SMD_07	0.93	1.01	1.05	SMH_10	-0.01	0.94	0.92
SMD_08	-1.56	0.94	0.88	SMH_11	0.06	1.17	<b>1.21</b>

**Notas:** Infit MNSQ = Ajuste interno a través de la media cuadrática no estandarizada

Outfit MNSQ = Ajuste externo a través de la media cuadrática no estandarizada

**Fuente:** elaboración propia

En la subescala de *Manejo de la Información*, de los 28 ítems que la conforman, seis ítems (SMB\_12, SMC\_08, SMC\_10, SME\_09, SME\_12, SMF\_13) presentan valores ligeramente por encima de 1.2, pero incluso estas cifras sólo están ligeramente por fuera del rango aceptable de 0.8 a 1.2, como puede verse en la Tabla 12.

**Tabla 12. Valores Infit y Outfit de la subescala *Manejo de la Información* del Excale 09 de Matemáticas**

Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ	Folio Ítem	Dificultad del ítem	Infit MNSQ	Outfit MNSQ
SMA_09	1.5	1.04	1.13	SME_09	0.79	1.13	<b>1.29</b>
SMA_10	0.26	1.04	1.07	SME_10	-1.07	0.95	0.92
SMA_11	1.08	1.07	1.11	SME_11	0.59	0.96	0.99
SMB_11	-0.7	0.93	0.9	SME_12	-0.87	0.84	<b>0.79</b>
SMB_12	1.1	1.09	<b>1.27</b>	SME_13	0.31	1.02	1.02
SMB_13	1.25	1.04	1.13	SMF_09	-0.92	0.91	0.87
SMC_04	-0.15	0.95	0.94	SMF_11	-0.04	0.97	0.98
SMC_07	0.5	1	1.03	SMF_12	0.07	1	1.02
SMC_08	-2.55	0.91	<b>0.7</b>	SMF_13	1.2	1.07	<b>1.25</b>
SMC_09	-0.12	0.91	0.89	SMG_11	0.17	0.98	0.98
SMC_10	0.78	1.16	<b>1.33</b>	SMG_12	-1.46	0.94	0.86
SMC_11	-0.4	0.95	0.92	SMH_06	0.09	1.04	1.08
SMC_12	-0.14	1.02	1.05	SMH_12	-0.47	0.98	0.98
SMD_11	0.19	1.11	1.18	SMH_13	-1	0.96	0.93

**Notas:** Infit MNSQ = Ajuste interno a través de la media cuadrática no estandarizada  
 Outfit MNSQ = Ajuste externo a través de la media cuadrática no estandarizada

**Fuente:** elaboración propia

En síntesis, la gran mayoría de los ítems de tienen propiedades psicométricas óptimas, y los pocos en los que no tienen niveles ideales de ajuste, su distorsión con el modelo de Rasch es muy baja, por lo que se concluye que estos reactivos cumplen con esta condición para continuar con los análisis DIF, que se presentan en los siguientes subapartados.

#### 4.2. Funcionamiento Diferencial de Ítems por Sexo

El DIF ocurre cuando dos sujetos con la misma habilidad que se desea medir, tienen diferentes probabilidades de responder correctamente a un mismo ítem en específico, ocasionada por su pertenencia a cierta subpoblación o grupos de clasificación (Boone *et al.*, 2014). Si los Excale 09 presentasen una alta magnitud



de DIF, sería un riesgo para la toma de decisiones a partir de sus resultados, especialmente, para formular políticas educativas que pretendan impactar a aquellos grupos de estudiantes donde, por el hecho de pertenecer a cierta subpoblación, tienen menos probabilidades de responder correctamente a ciertos ítems.

El análisis DIF por sexo, se realizó para cada una de las tres subescalas que mide el Excale 09 de Matemáticas, a saber: sentido numérico y pensamiento algebraico (SNPA); forma, espacio y medida (FEM); y manejo de la información (MI).

El análisis DIF se hizo bajo dos modalidades: identificar la presencia de DIF Uniforme y No Uniforme. Como se señaló en el capítulo metodológico de esta tesis, un DIF Uniforme consiste en la presencia de un funcionamiento diferencial a lo largo de todos los niveles de habilidad de los alumnos; en cambio, un DIF No Uniforme (NUDIF) se presenta cuando en grupos de alumnos con baja habilidad, una de las subpoblaciones clasificatorias (por ejemplo, los hombres) tienen más probabilidades de responder correctamente a un ítem, y en grupos de alumnos con alta habilidad, otra de las subpoblaciones clasificatorias (por ejemplo, las mujeres) son quienes tienen más posibilidades de contestar correctamente al mismo ítem.

#### **4.2.1. DIF Uniforme por Sexo**

En la Figura 2 se ilustran los resultados del análisis DIF que se realizó a la subescala de Matemáticas llamada SNPA, realizados mediante el software Winsteps. En el eje de las  $X$  se ubicó a cada uno de los ítems mientras que en el

eje de las Y se graficó la diferencia entre el nivel de dificultad de un reactivo ( $d_i$ ) y el nivel de dificultad promedio del ítem a los hombres ( $d_{i-h}$ ) y para las mujeres ( $d_{i-m}$ ). Así, un ítem que no presentase DIF es aquel en el que los valores graficados para hombres y mujeres se ubicasen en el cero. Cuanta más distancia haya entre un valor y el cero indica un mayor funcionamiento diferencial para la subpoblación analizada. Valores positivos muestran mayor dificultad del ítem para uno de los grupos de que se trate; por el contrario, valores negativos son indicativos de mayor facilidad para alguna de las dos subpoblaciones de análisis.

Aun cuando se señalarán los reactivos puntuales que presentaron DIF, para que el lector disponga de una visión gráfica de los ítems con un funcionamiento diferencial, a las figuras se agregan cuatro líneas de referencia que sirven para identificar el grado de DIF de los ítems. Las líneas situadas en los  $|0.43|$  lógitos sirven para distinguir entre reactivos con DIF moderado y aquéllos cuyo funcionamiento diferencial es ligero o poco significativo; por su parte, las líneas colocadas a los  $|0.64|$  lógitos permiten diferenciar entre ítems con DIF severo y moderado. Adviértase que la escala del eje de las “Y” es distinta según los valores que se presenten en cada caso.

La evidencia empírica muestra que ninguno de los cuarenta ítems de la subescala SNPA presenta DIF entre hombres y mujeres, ni siquiera moderado. El ítem SMD\_10, que fue el que presentó mayor DIF, fue de 0.42 lógitos a favor de las mujeres ( $p=0.01$ ;  $T=77.89$ )<sup>8</sup> lo que todavía se considera como DIF ligero. En el Anexo 8 se presenta, para cada ítem, la magnitud puntual de DIF, el nivel de

---

<sup>8</sup> Para aligerar la lectura, en adelante se omitirán los valores de p y T cuando  $p<0.05$  y  $T\geq 2$ .



favorece a los hombres que son el grupo de referencia y el signo positivo "+" representa que es más fácil para las mujeres que son el grupo focal.

Los resultados con TAM confirmaron que ninguno de los cuarenta ítems de la subescala *Sentido Numérico y Pensamiento Algebraico* presenta DIF uniforme severo, lo cual es un hallazgo muy positivo. Mediante el método de estimación EMMV se identificaron dos ítems (SMB\_01 y SMD\_10) con DIF pero sólo fue moderado, el resto presentó DIF insignificante.

**Tabla 13. Nivel y sentido del DIF mediante paquetería TAM en la subescala Sentido Numérico y Pensamiento Algebraico, desglosado por Sexo**

Ítem	<i>d</i> Hom	<i>d</i> Muj	Dife-rencia entre grupos	Nivel y Senti-do del DIF	Ítem	<i>d</i> Hom	<i>d</i> Muj	Dife-rencia entre grupos	Nivel y Senti-do del DIF
SMA_01	0.71	0.66	0.05	A+	SME_03	0.27	-0.04	0.31	A+ *
SMA_02	-1.18	-0.82	-0.36	A- *	SMF_01	-0.15	-0.28	0.13	A+ *
SMA_05	-0.06	0.03	-0.08	A- *	SMF_02	-1.15	-1.47	0.32	A+ *
SMA_08	-1.60	-1.57	-0.03	A-	SMF_03	0.49	0.56	-0.06	A-
SMA_12	-0.12	-0.14	0.02	A+	SMF_04	-0.19	-0.14	-0.05	A-
SMB_01	-0.11	0.36	-0.47	B- *	SMF_05	-0.10	-0.11	0.02	A+
SMB_02	0.36	0.37	-0.02	A-	SMG_01	1.09	1.00	0.09	A+ *
SMB_03	-0.10	0.00	-0.10	A- *	SMG_02	0.00	-0.11	0.11	A+ *
SMB_04	0.87	0.79	0.08	A+ *	SMG_03	0.41	0.27	0.14	A+ *
SMB_05	-1.33	-1.41	0.08	A+ *	SMG_04	-1.03	-0.63	-0.40	A- *
SMC_01	-0.08	-0.11	0.03	A+	SMG_05	-0.02	0.09	-0.11	A- *
SMC_02	-0.28	-0.40	0.13	A+ *	SMG_08	0.04	0.18	-0.14	A- *
SMC_03	0.33	0.43	-0.10	A- *	SMG_09	-0.75	-0.64	-0.12	A- *
SMC_05	1.46	1.68	-0.22	A- *	SMG_10	-0.18	-0.21	0.02	A+
SMC_06	0.04	0.03	0.01	A+	SMH_01	-0.70	-0.78	0.08	A+ *
SMD_01	0.90	0.79	0.10	A+ *	SMH_02	0.36	0.37	-0.01	A-
SMD_10	-0.34	-0.76	0.43	B+ *	SMH_03	0.06	0.13	-0.07	A- *
SMD_12	0.60	0.67	-0.07	A- *	SMH_04	1.21	1.30	-0.10	A- *
SME_01	0.11	0.11	0.00	A+	SMH_05	-0.10	-0.31	0.21	A+ *
SME_02	-0.42	-0.56	0.14	A+ *	SMH_07	0.66	0.68	-0.02	A-

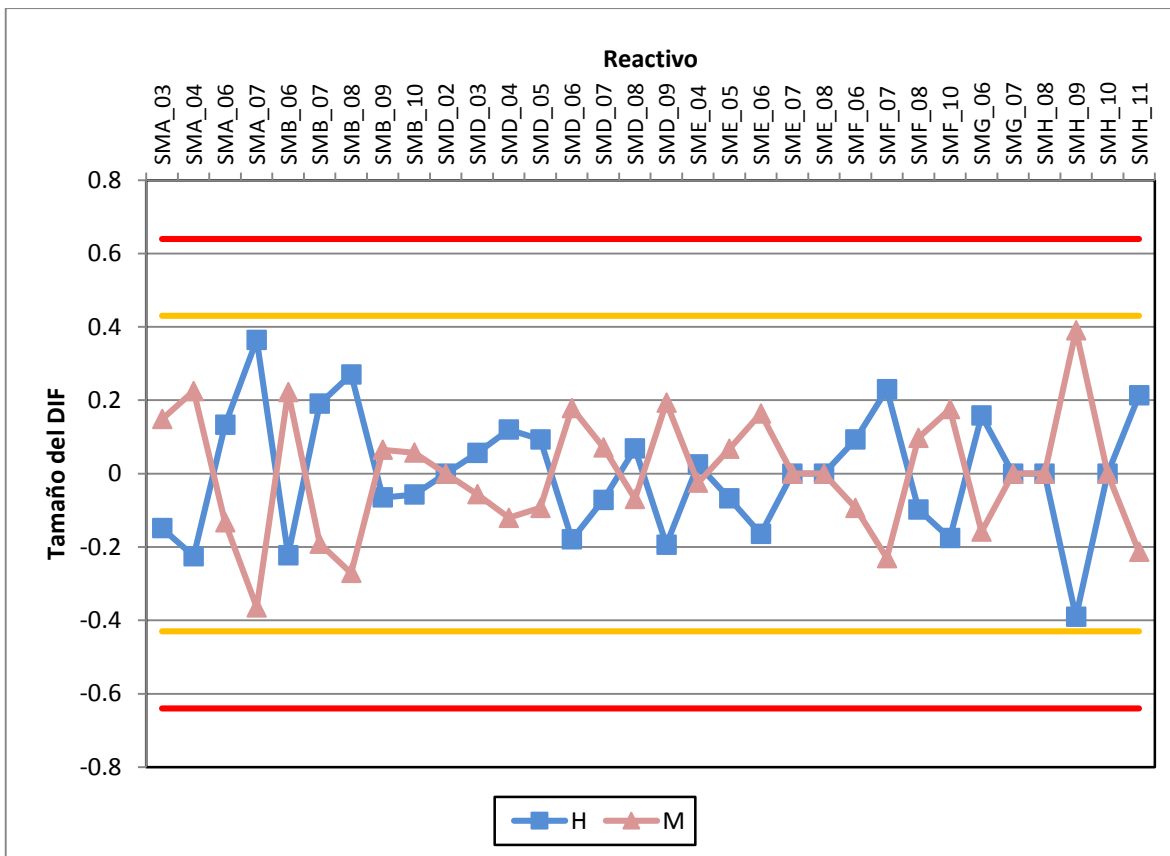
**Notas:** A=DIF insignificante; B= DIF moderado; C=DIF severo

\* = DIF estadísticamente significativo al 0.005

- = DIF favorece a grupo de referencia; + = DIF favorece grupo focal

**Fuente:** elaboración propia

En lo que corresponde a la subescala *Forma, Espacio y Medida*, lo que muestra la evidencia es que cada uno de sus 32 ítems está libre de DIF uniforme de forma significativa por sexo, lo cual implica que tanto hombres como mujeres tienen las mismas posibilidades de responder correctamente a los ítems de esta subescala, sólo en función de la habilidad matemática sobre el dominio FEM (ver Figura 3). Aun los reactivos con más DIF por sexo, tales como el SMA\_07 y SMH\_09, su magnitud no supera los 0.39 lógitos, lo que indica DIF insignificante (ver Anexo 9).



**Figura 3** Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala *Forma, Espacio y Medida* del Excale 09 de Matemáticas en 2012, desglosado por Sexo

Los resultados con TAM confirman que ninguno de los 32 ítems de la subescala FEM presenta DIF uniforme severo por sexo, ni siquiera DIF moderado. Con el método de estimación EMMV, ninguno de los ítems tiene una diferencia en los niveles de dificultad que supere los 0.33 lógitos, lo que los sitúa como ítems con DIF insignificante (ver Tabla 14).

**Tabla 14. Nivel y sentido del DIF mediante paquetería TAM en la subescala Forma, Espacio y Medida, desglosado por Sexo**

Ítem	Nivel Dificultad Hombres	Nivel Dificultad Mujeres	Diferencia entre grupos	Nivel y Sentido del DIF
SMA_03	0.10	0.22	-0.12	A- *
SMA_04	-0.17	-0.03	-0.14	A- *
SMA_06	-0.03	-0.12	0.09	A+ *
SMA_07	0.11	-0.22	0.33	A+ *
SMB_06	-0.20	-0.08	-0.13	A- *
SMB_07	-0.95	-1.22	0.26	A+ *
SMB_08	0.72	0.45	0.27	A+ *
SMB_09	-0.34	-0.36	0.02	A+
SMB_10	0.16	0.23	-0.08	A- *
SMD_02	1.32	1.37	-0.06	A-
SMD_03	-0.81	-0.92	0.11	A+ *
SMD_04	0.69	0.50	0.19	A+ *
SMD_05	-0.30	-0.36	0.06	A+
SMD_06	0.82	0.96	-0.13	A- *
SMD_07	0.75	0.88	-0.13	A- *
SMD_08	-1.35	-1.47	0.11	A+ *
SMD_09	-0.95	-0.89	-0.06	A-
SME_04	0.80	0.83	-0.03	A-
SME_05	-0.69	-0.51	-0.18	A- *
SME_06	-0.95	-0.78	-0.17	A- *
SME_07	0.22	0.18	0.04	A+
SME_08	-0.06	-0.09	0.03	A+
SMF_06	0.23	0.11	0.12	A+ *
SMF_07	-0.21	-0.37	0.17	A+ *
SMF_08	1.13	1.17	-0.04	A-
SMF_10	-1.08	-0.87	-0.21	A- *
SMG_06	-0.93	-1.00	0.08	A+ *
SMG_07	2.26	2.35	-0.10	A- *
SMH_08	0.58	0.63	-0.05	A-
SMH_09	-0.91	-0.57	-0.33	A- *
SMH_10	-0.03	0.04	-0.07	A- *
SMH_11	0.07	-0.07	0.15	A+ *

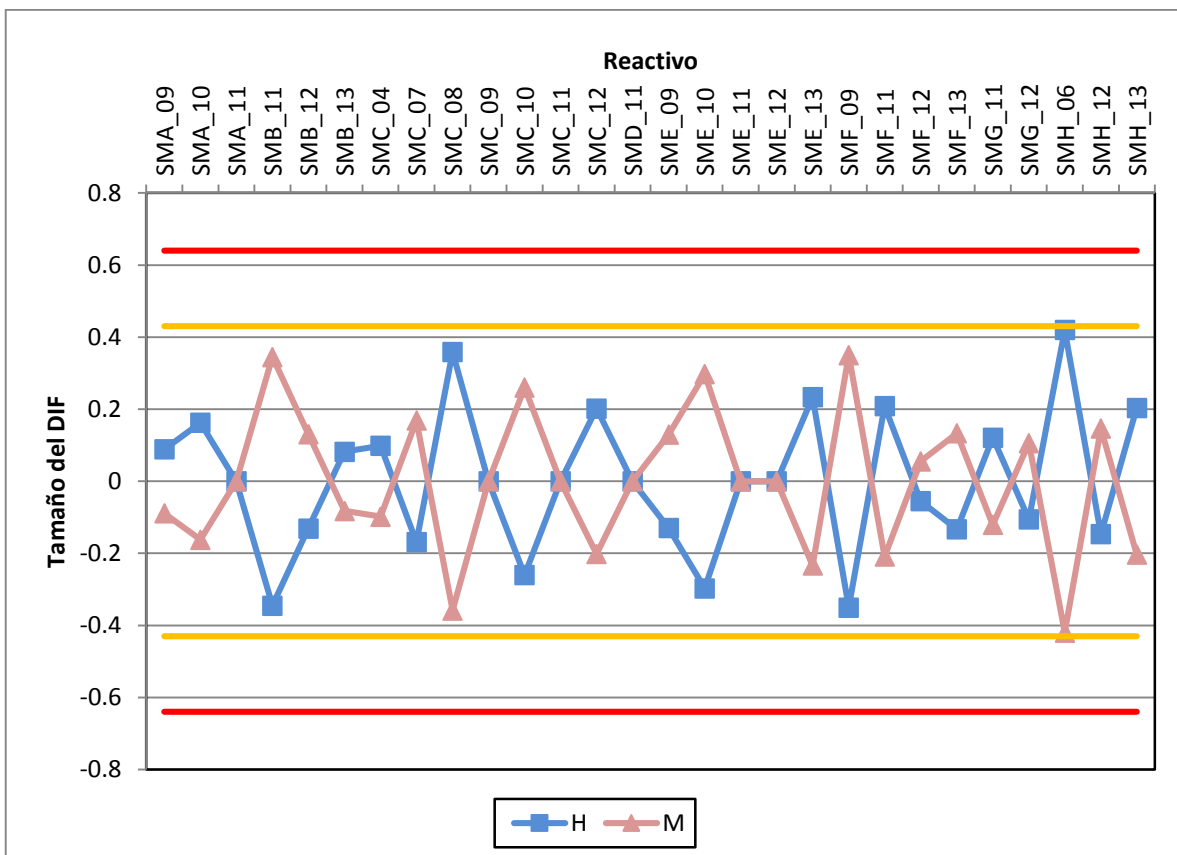
**Notas:** A=DIF insignificante; B= DIF moderado; C=DIF severo

\* = DIF estadísticamente significativo al 0.005

- = DIF favorece a grupo de referencia; + = DIF favorece grupo focal

**Fuente:** elaboración propia

En lo que referente a la subescala Manejo de la Información, los datos muestran que los 28 ítems que la conforman están libres de DIF uniforme severo por sexo (ver Figura 4), lo cual implica que tanto hombres como mujeres tienen las mismas posibilidades de responder correctamente a los ítems de esta subescala, y que más bien, lo que diferencia los puntajes que cada alumno obtiene es su habilidad matemática sobre el dominio MI. El ítem SMH\_06, que es el que presenta mayor DIF, es de 0.42 lógitos, valor que aun se considera aceptable (ver Anexo 10).



**Figura 4. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Sexo**



Los resultados con TAM concurren en que los 28 ítems de la subescala MI están libres de DIF uniforme severo por sexo. Utilizando el método de estimación EMMV, al igual que con el software Rasch, ninguno de los ítems tiene una diferencia en los niveles de dificultad que supere los 0.40 lógitos, lo que los sitúa como ítems con DIF insignificante (ver Tabla 15).

**Tabla 15. Nivel y sentido del DIF mediante paquetería TAM en la subescala Manejo de la Información, desglosado por Sexo**

Ítem	Nivel Dificultad Hombres	Nivel Dificultad Mujeres	Diferencia entre grupos	Nivel y Sentido del DIF
SMA_09	1.31	1.32	-0.01	A-
SMA_10	0.29	0.07	0.22	A+ *
SMA_11	1.02	0.92	0.10	A+ *
SMB_11	-0.83	-0.50	-0.33	A- *
SMB_12	0.92	0.95	-0.03	A-
SMB_13	1.14	1.19	-0.06	A-
SMC_04	-0.04	-0.12	0.09	A+ *
SMC_07	0.38	0.58	-0.20	A- *
SMC_08	-2.10	-2.50	0.40	A+ *
SMC_09	-0.06	-0.14	0.08	A+ *
SMC_10	0.57	0.83	-0.26	A- *
SMC_11	-0.31	-0.30	0.00	A-
SMC_12	0.03	-0.19	0.22	A+ *
SMD_11	0.23	0.17	0.06	A+
SME_09	0.60	0.78	-0.18	A- *
SME_10	-1.03	-0.81	-0.22	A- *
SME_11	0.48	0.61	-0.13	A- *
SME_12	-0.73	-0.70	-0.03	A-
SME_13	0.42	0.23	0.19	A+ *
SMF_09	-1.01	-0.69	-0.32	A- *
SMF_11	0.08	-0.21	0.28	A+ *
SMF_12	-0.05	0.04	-0.09	A- *
SMF_13	1.00	1.18	-0.19	A- *
SMG_11	0.16	0.12	0.04	A+
SMG_12	-1.25	-1.25	-0.01	A-
SMH_06	0.17	-0.13	0.31	A+ *
SMH_12	-0.49	-0.39	-0.10	A- *
SMH_13	-0.90	-1.06	0.16	A+ *

**Notas:** A=DIF insignificante; B= DIF moderado; C=DIF severo

\* = DIF estadísticamente significativo al 0.005

- = DIF favorece a grupo de referencia; + = DIF favorece grupo focal

**Fuente:** elaboración propia

#### 4.2.2. DIF No Uniforme por Sexo

La presencia de NUDIF ocurre cuando al desagregar a la población de estudiantes de acuerdo a niveles de habilidad, en algunos de esos niveles los ítems tienen más probabilidad de ser respondidos correctamente por alguno de los grupos de interés (por ejemplo, los hombres) y, en un nivel de habilidad distinto al anterior, otro grupo de interés (por ejemplo, las mujeres) es quien tiene más probabilidades de responderlo de forma correcta. Es decir, la detección de NUDIF equivale a la realización de varios análisis DIF desagregando a los sustentantes según la habilidad evaluada.

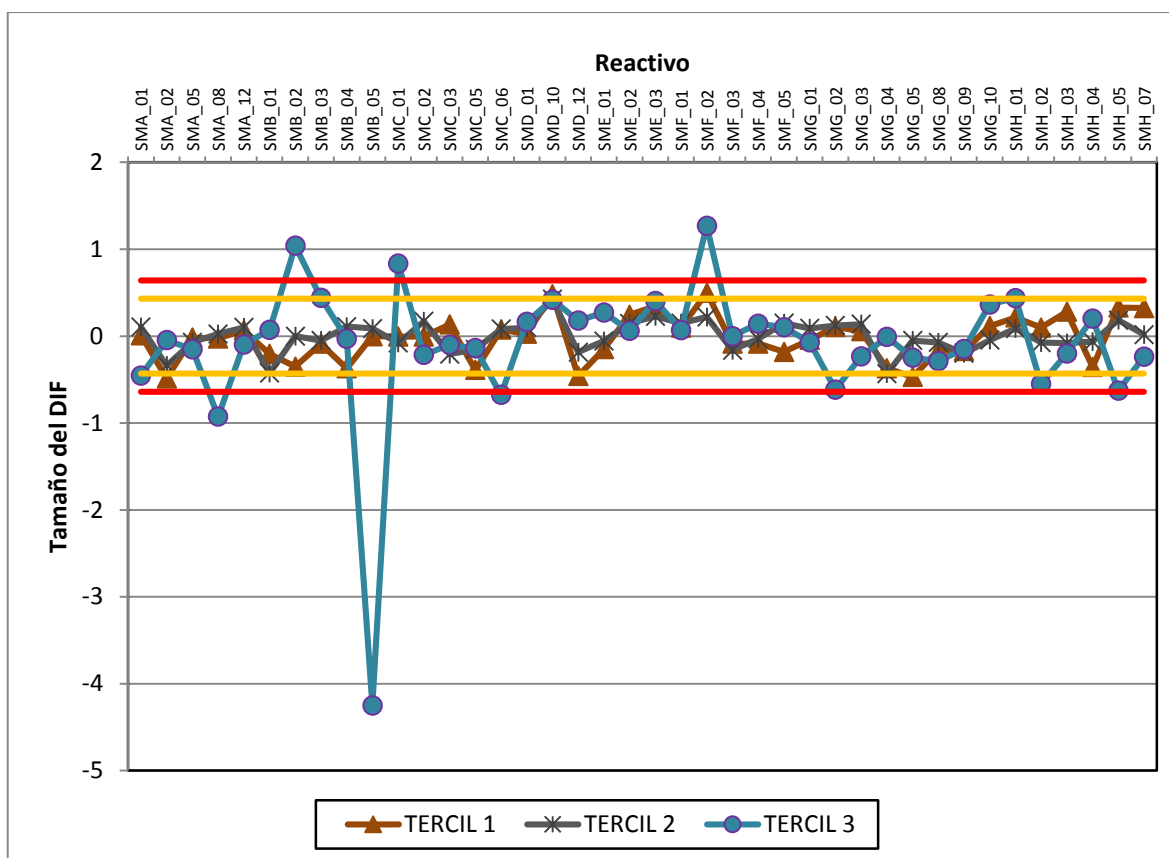
En esta investigación, para facilitar la presentación de los resultados, se decidió dividir a los sustentantes del Excale 09 en terciles según su habilidad matemática. De esta manera se analizó si los ítems presentaban un funcionamiento que favoreciera a hombres o mujeres a lo largo de cada uno de los tres niveles de habilidad evaluada de los alumnos, a diferencia de los análisis previos llevados a cabo para identificar DIF uniforme, donde no se requería desagregar a los estudiantes según su nivel de habilidad matemática.

En la Figura 5 se presentan los análisis NUDIF llevados a cabo entre hombres y mujeres con cada uno de los terciles en que se subdividió a los estudiantes según su habilidad matemática en la subescala SNPA. En el eje de las X se ubicó a cada uno de los ítems mientras que en el eje de las Y se graficó la diferencia entre el nivel de dificultad de un ítem para hombres ( $d_{i-h}$ ) y mujeres ( $d_{i-m}$ ) de cada uno de los terciles definidos en función de su habilidad matemática. Valores positivos

indican que el ítem es más fácil para las mujeres; por el contrario, valores negativos son indicativos de que el ítem favorece a los hombres.

La evidencia empírica muestra que ninguno de los cuarenta reactivos de la subescala SNPA presenta NUDIF por sexo, recordando que para que haya se requiere que el mismo ítem presente DIF-Contrast severo a favor de un grupo en un tercil determinado, y a favor del otro grupo en un tercil distinto. Sin embargo, el análisis al interior de cada tercil (que llamaremos DIF parcial) permitió encontrar algunos ítems que presentan DIF parcial moderado (ver Figura 5).

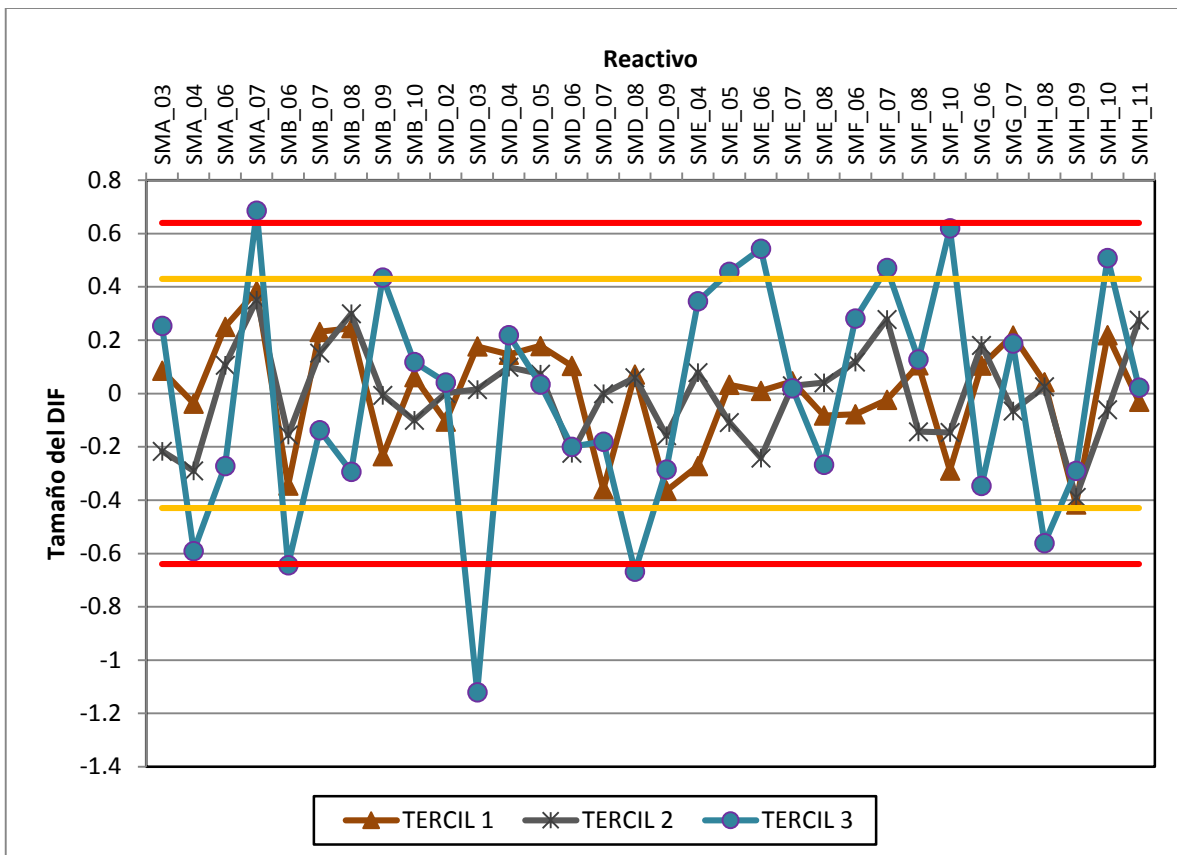
En el primer tercil (33.3% de los estudiantes con menor desempeño en Matemáticas), cinco reactivos tienen DIF parcial moderado, de los cuales tres son favorables para los hombres (SMA\_02, SMD\_12 y SMG\_05) y dos ítems para las mujeres (SMD\_10 y SMF\_02). En el segundo tercil, se identificaron dos ítems con DIF parcial moderado (SMB\_01 y SMG\_04), ambos con mayor probabilidad para que sean respondidos correctamente por los hombres. En el tercil superior (33.3% de estudiantes con mayores habilidades matemáticas) se identificaron 12 reactivos con presencia de DIF parcial, de los cuales la mitad tiene DIF severo, mayor a 0.64 lógitos. Comparando a la subpoblación a la que favorecen, siete ítems son favorables a los hombres (SMA\_01, SMA\_08, SMB\_05, SMC\_06, SMG\_02, SMH\_02 y SMH\_05) y cinco a las mujeres (SMB\_02, SMB\_03, SMC\_01, SMF\_02 y SMH\_01). En el Anexo 11 se presentan los datos puntuales con los resultados sobre el NUDIF entre hombres y mujeres.



**Figura 5. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Sentido Numérico y Pensamiento Algebraico* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo**

En la subescala FEM la información disponible indica que ninguno de los 32 ítems que la conforman presenta NUDIF severo por sexo (ver Figura 6). De hecho, al llevar a cabo un análisis al interior de cada tercil de habilidad (DIF parcial), se halló que en los primeros dos terciles todos los reactivos están libres de DIF parcial severo. En el tercil superior de habilidad se identificó una docena de ítems con presencia de DIF parcial, de los cuales tres tienen DIF severo o mayor a 0.64 lógitos. Comparando a la subpoblación a la que favorecen, cinco reactivos son favorables a los hombres (SMA\_04, SMB\_06, SMD\_03, SMD\_08 y SMH\_08) y los

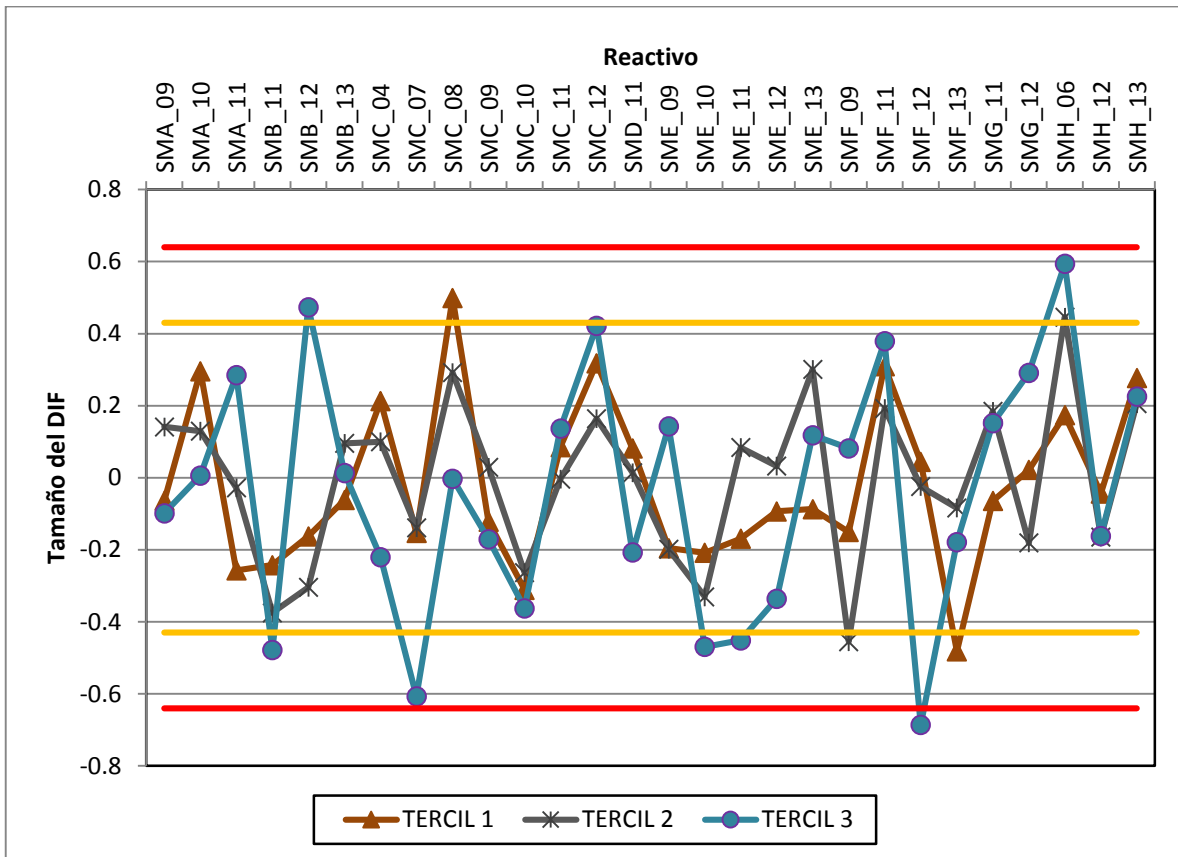
siete restantes con las mujeres (SMA\_07, SMB\_09, SME\_05, SME\_06, SMF\_07, SMF\_10 y SMH\_10) (ver Anexo 12).



**Figura 6 Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Forma, Espacio y Medida* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo**

El análisis NUDIF de la subescala MI indica que sus 28 ítems están libres de NUDIF severo por sexo (ver Figura 7); sin embargo, se identificaron algunos reactivos con DIF parcial. Así, en el tercil inferior los ítems SMC\_08 y SMF\_13 tienen un DIF parcial moderado que favorece a las mujeres. En el segundo tercil

se identificó un par de ítems (SMF\_09 y SMH\_06) con DIF parcial moderado que favorece a los hombres. En el tercil superior se diagnosticaron siete ítems con DIF parcial, de los cuales cinco otorgan mayores posibilidades de responderse correctamente por los hombres (SMB\_11, SMC\_07, SME\_10, SME\_11 y SMF\_12) y un par de ítems (SMB\_12 y SMH\_06) que son más fáciles de responder por las mujeres (ver Anexo 13).



**Figura 7. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Manejo de la Información* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Sexo**

### **4.3. Funcionamiento Diferencial de Ítems por NSE**

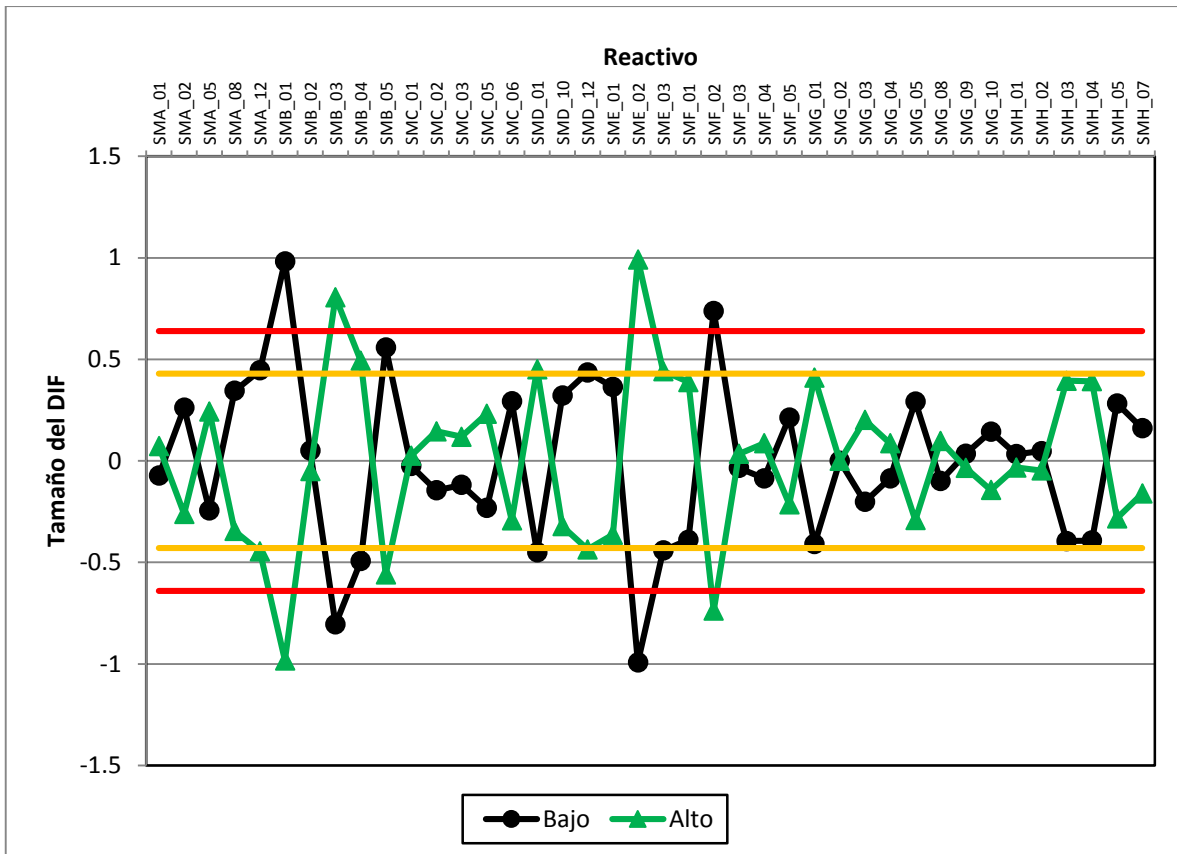
El análisis DIF por NSE, al igual que con el sexo, se realizó para cada una de las tres subescalas que mide el Excale 09 de Matemáticas: SNPA, FEM y MI. De igual forma, el análisis DIF se hizo bajo dos modalidades: identificar la presencia de DIF y NUDIF.

Como ya se señaló en el apartado metodológico de esta tesis, a los alumnos se les clasificó en tres niveles socioeconómicos (Bajo, Medio y Alto) tomando en consideración información sobre algunos bienes de los que disponían en sus hogares (automóviles, teléfono, computadora), máximo nivel educativo de sus padres, índice de hacinamiento, entre otros.

#### **4.3.1. DIF Uniforme por NSE**

Los resultados de los análisis DIF por NSE llevados a cabo con los ítems de la subescala SNPA, permitieron saber que diez de los cuarenta ítems que la conforman presentan funcionamiento diferencial, seis de ellos de forma moderada y cuatro de forma severa (ver Figura 8). De los ítems con DIF moderado, tres de ellos tienen un funcionamiento diferencial que favorece a los alumnos de NSE Alto (SMA\_12, SMB\_05, y SMD\_12), y la otra mitad a estudiantes de NSE Bajo (SMB\_04, SMD\_01 y SME\_03); de los cuatro ítems con DIF severo, dos de ellos favorecen a estudiantes de NSE Alto (SMB\_01 y SMF\_02) mientras que la otra mitad a alumnos que provienen de familias con NSE Bajo (SMB\_03 y SME\_02) (ver Anexo 14).

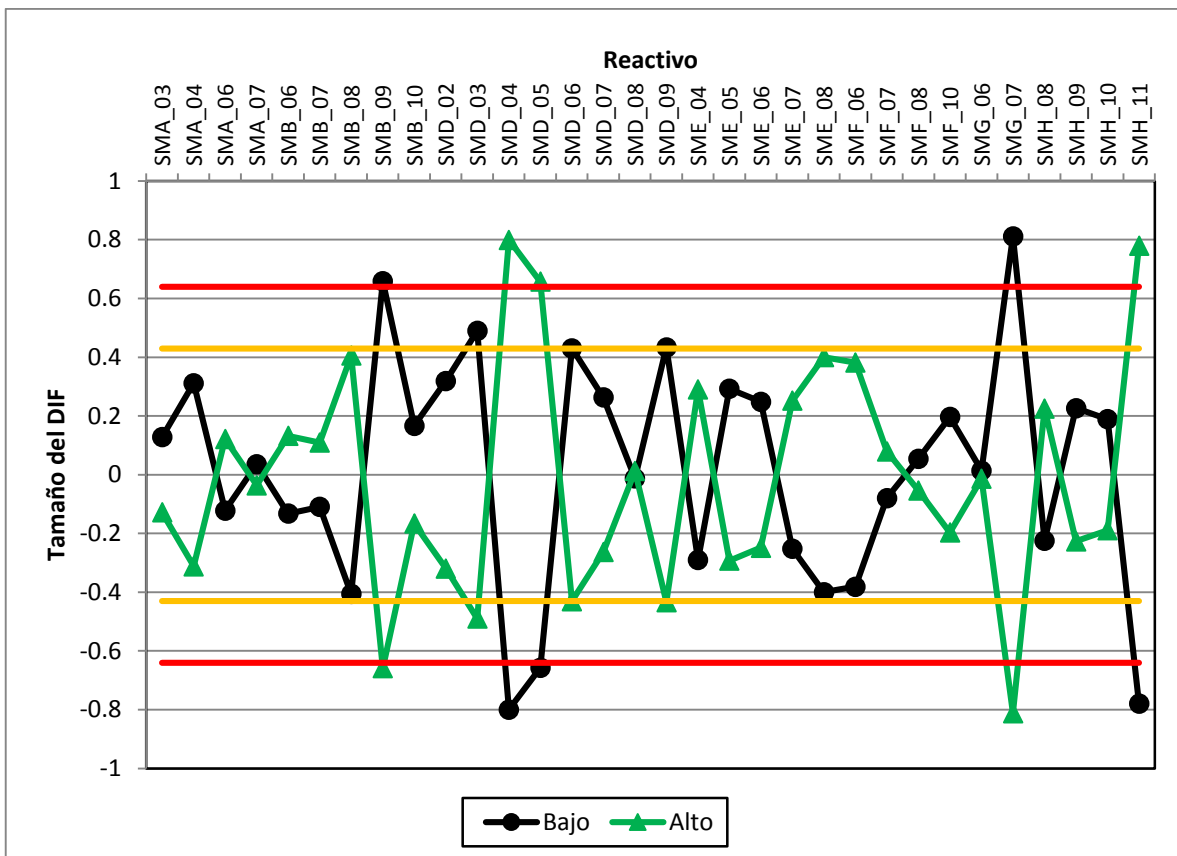




**Figura 8** Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala *Sentido Numérico y Pensamiento Algebraico* del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico

En la subescala FEM se encontró que si bien la gran mayoría de sus 32 ítems no presentan un comportamiento en los que haya una sospecha de sesgo, cinco reactivos exhiben un DIF severo entre alumnos con diferentes NSE (ver Figura 9); los ítems SMD\_04, SMD\_05 y SMH\_11 tienen un DIF a favor de los estudiantes con NSE Bajo y los ítems SMB\_09 y SMG\_07 muestran un DIF que favorece a los alumnos con NSE Alto. Asimismo, la evidencia muestra que tres reactivos

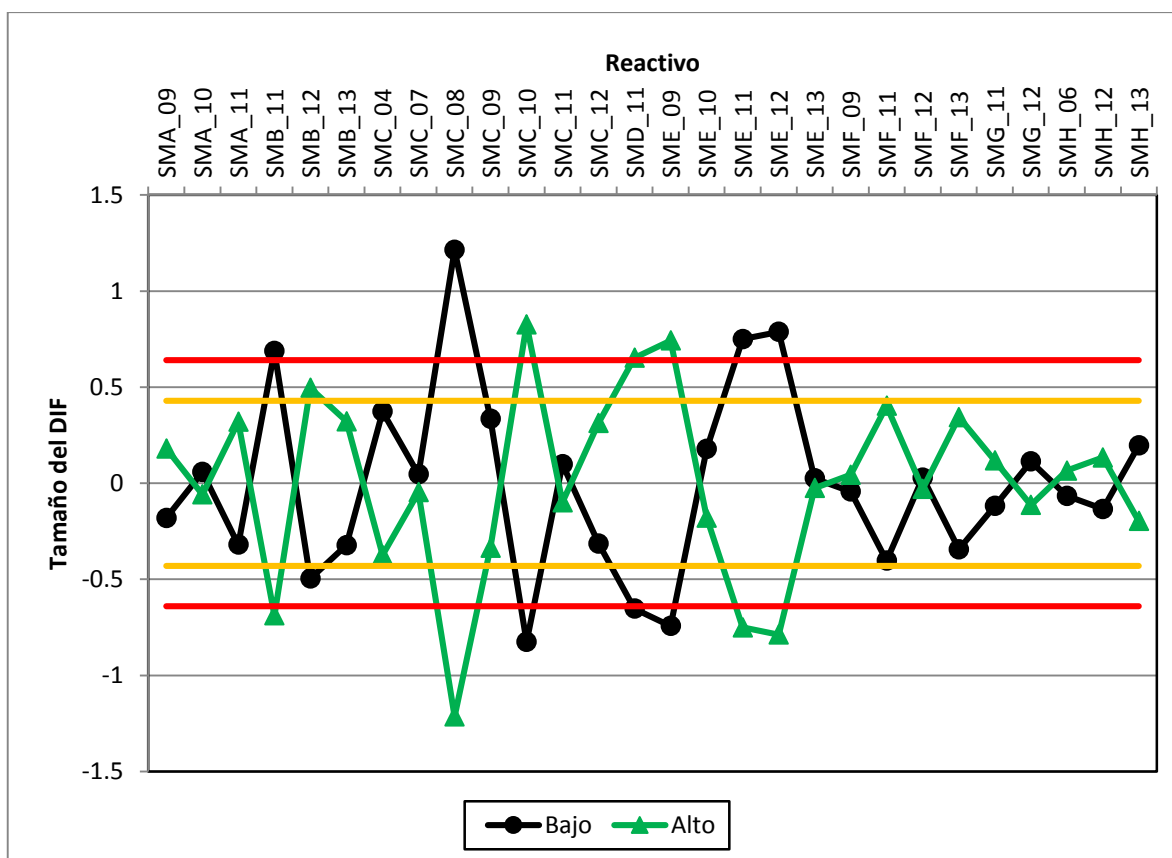
adicionales mostraron un DIF moderado favorable hacia los alumnos con NSE Alto, a saber: SMD\_03, SMD\_06 y SMD\_09 (ver Anexo 15).



**Figura 9** Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala *Forma, Espacio y Medida* del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico

En la subescala MI se halló que la mayoría de los 28 ítems no presentan sesgo (ver Figura 10); sin embargo, se identificaron siete ítems con DIF severo, de los cuales cuatro favorecen a los alumnos con NSE Alto (SMB\_11, SMC\_08, SME\_11 y SME\_12), y los tres restantes tienen un DIF que da ventaja a estudiantes de NSE Bajo (SMC\_10, SMD\_11 y SME\_09). Además, se encontró que el ítem

SMB\_12 tiene un DIF moderado a favor de los estudiantes con NSE Bajo (ver Anexo 16).



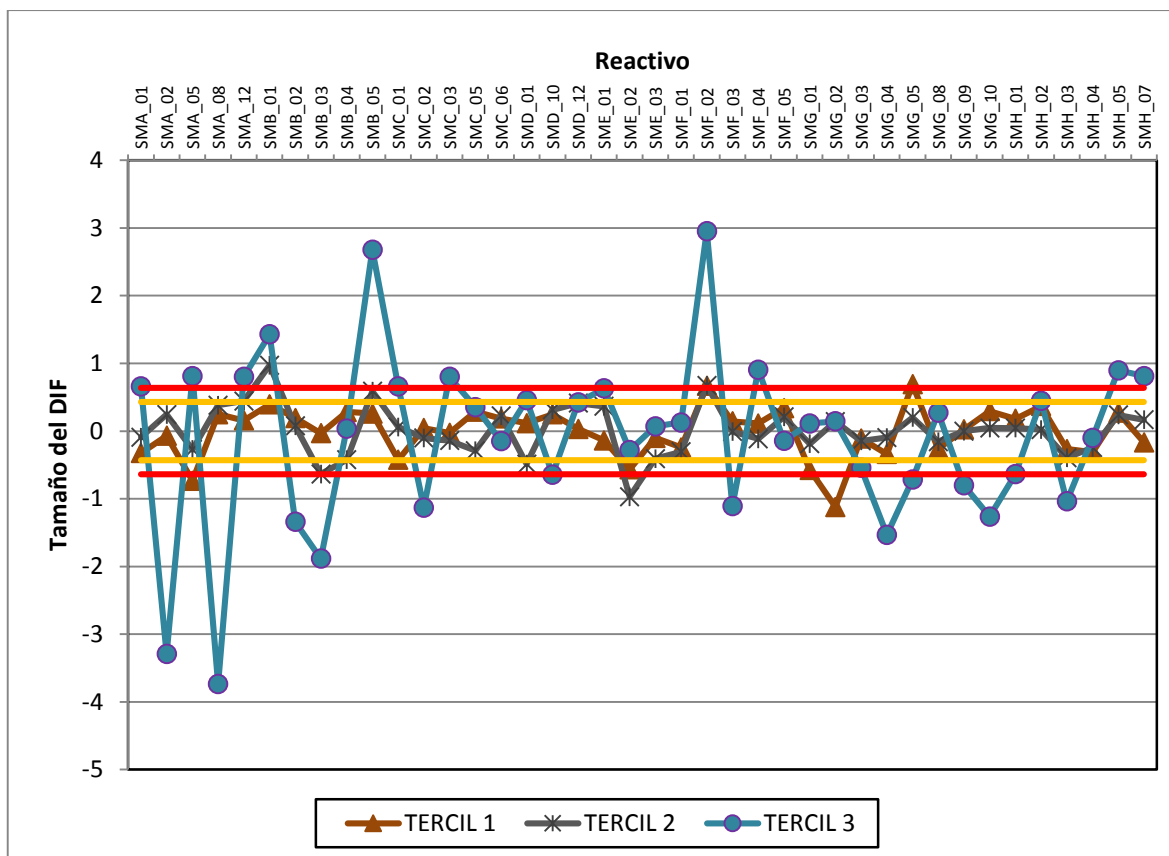
**Figura 10. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala *Manejo de la Información* del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

#### 4.3.2. DIF No Uniforme por NSE

Como se hizo con el análisis NUDIF por sexo, se dividió a los sustentantes del Excale 09 en terciles según su habilidad matemática. De esta manera, se analizó si los ítems presentaban un funcionamiento que favoreciera a alumnos con NSE alto o bajo a lo largo de cada uno de los tres niveles en que se dividió la habilidad matemática de los alumnos.

En la subescala SNPA, al llevar a cabo los análisis al interior de cada tercil de habilidad (lo que se identificó como DIF parcial), la evidencia señala que entre el subgrupo del 33% de los alumnos con menos rendimiento en Matemáticas, hay cuatro ítems (SMA\_05, SMG\_02, SMF\_02 y SMG\_05) con DIF parcial de al menos 0.64 lógitos, de los cuales los dos primeros son más fáciles para los alumnos con NSE Bajo y los dos últimos ítems son más fáciles para estudiantes con NSE Alto (ver Figura 11). Vale recordar que aunque esta condición es necesaria no es suficiente para diagnosticar NUDIF, sino que se requiere que el DIF sea en uno de los terciles favorable hacia los alumnos de NSE Alto y en otro de los terciles hacia los estudiantes de NSE Bajo.

En el Tercil 2 de habilidad matemática, se diagnosticaron tres ítems con DIF parcial severo. Los ítems SMB\_01 y SMF\_02 son favorables para alumnos con NSE Alto y el reactivo SME\_02 es más fácil para estudiantes con NSE Bajo. Dentro del subgrupo conformado por la tercera parte de los alumnos de 3° de secundaria con los puntajes más altos en Matemáticas se hallaron once ítems con DIF parcial severo a favor de estudiantes con NSE Alto (SMA\_01, SMA\_05, SMA\_12, SMB\_01, SMB\_05, SMC\_01, SMC\_03, SMF\_02, SMF\_04, SMH\_05 y SMH\_07) y otros once ítems que son más fáciles para alumnos con NSE Bajo (SMA\_02, SMA\_08, SMB\_02, SMB\_03, SMC\_02, SMF\_03, SMG\_04, SMG\_05, SMG\_09, SMG\_10 y SMH\_03). En la misma Figura 11 podrá advertirse que la gran mayoría de los DIF severos ocurren en el tercil superior de habilidad matemática, es decir, entre los alumnos que suelen conseguir los puntajes más altos en el Excale 09 de Matemáticas (ver Anexo 17).

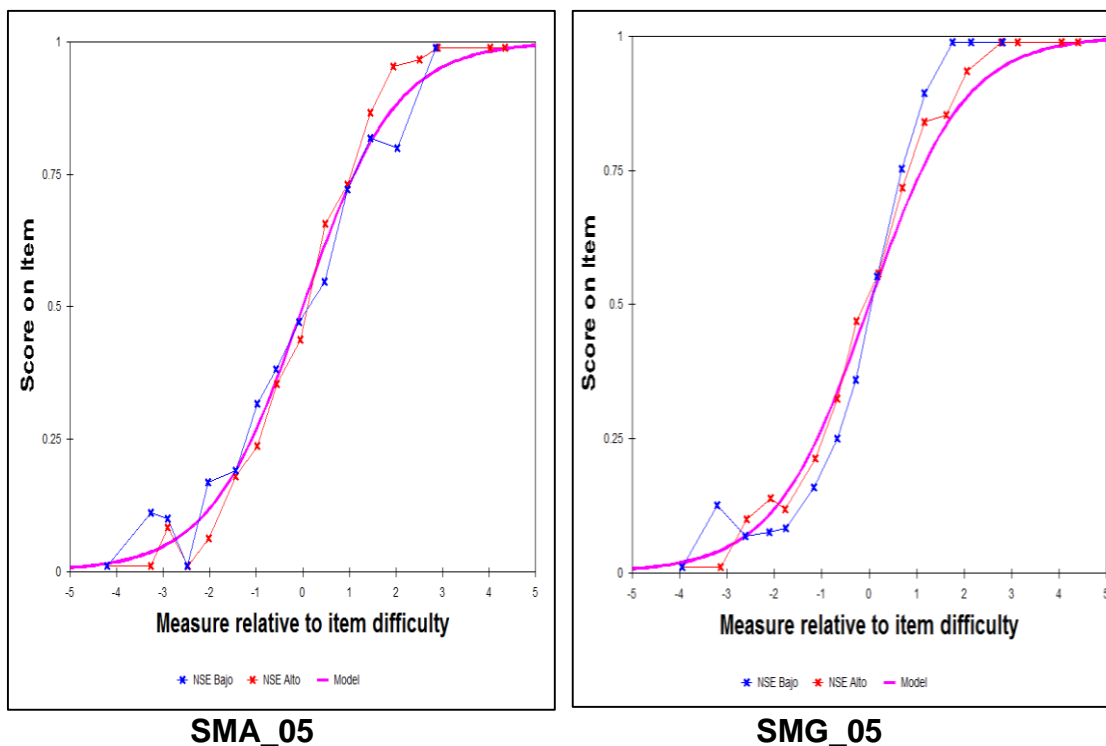


**Figura 11. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Sentido Numérico y Pensamiento Algebraico* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico**

En resumen, para la subescala SNPA se identificaron dos ítems como los únicos que cumplen el doble requisito para considerar que tiene NUDIF por NSE: SMA\_05 y SMG\_05. Debe subrayarse que estos ítems son los únicos en tal situación en las tres subescalas.

En el primero de ellos, dentro del subgrupo de estudiantes del primer tercil de habilidad matemática, los alumnos que provienen de un NSE Bajo son quienes tienen más probabilidades de responder correctamente al ítem; en cambio, dentro del subgrupo de alumnos con los más altos puntajes en la prueba, los estudiantes con NSE alto son los que tienen más probabilidades de acertar la respuesta

correcta. En el ítem SMG\_05 ocurre lo opuesto: entre los estudiantes del tercil más bajo de la habilidad, los que provienen de hogares con NSE Alto son los más propensos a resolver correctamente el reactivo; en cambio, al centrar la atención en el subgrupo que obtuvo los más altos puntajes, son los estudiantes de NSE Bajo los que tienen más probabilidad de resolverlo acertadamente (ver Figura 12).



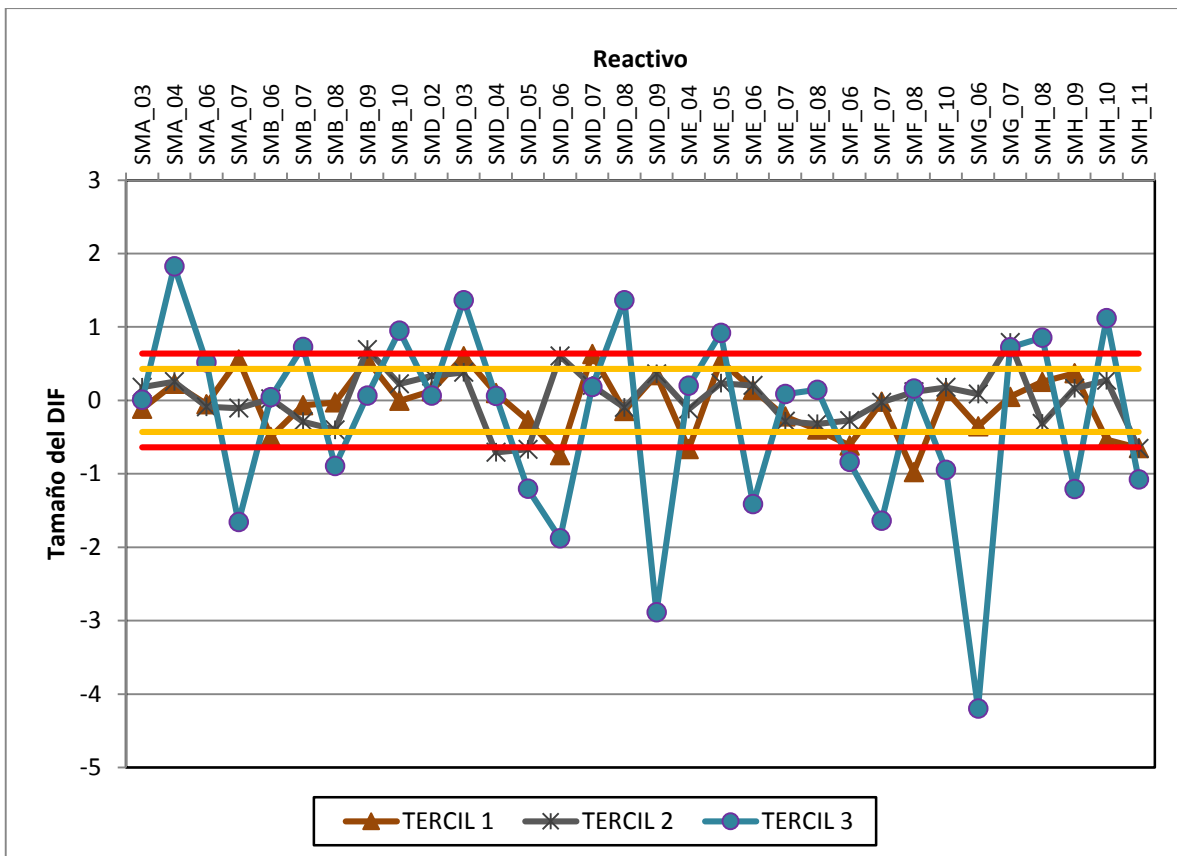
Fuente: elaboración propia.

**Figura 12. Curvas del Funcionamiento Diferencial de Ítems No Uniforme, según el Nivel Socioeconómico de los alumnos, de ítems SMA\_05 y SMG\_05**

En la subescala FEM ninguno de los 32 ítems que la conforman presentan NUDIF en grado severo, por NSE (ver Figura 13). No obstante, al realizar DIF parciales al interior de cada uno de los terciles del continuo de la escala de habilidad

matemática, la información disponible indica que en primer tercil (33% de los alumnos con menor puntaje en la prueba) hay cuatro ítems con DIF parcial severo a favor de alumnos de NSE Bajo: SMD\_06, SME\_04, SMF\_08 y SMH\_11.

Dentro del segundo tercil de habilidad matemática, se identificaron dos ítems (SMB\_09 y SMG\_07) con DIF parcial severo a favor de estudiantes con NSE Alto y tres que son más fáciles para los alumnos con NSE Bajo: SMD\_04, SMD\_05 y SMH\_11. En el tercil superior se encontraron nueve ítems (SMA\_04, SMB\_07, SMB\_10, SMD\_03, SMD\_08, SME\_05, SMG\_07, SMH\_08 y SMH\_10) con DIF parcial severo que son más fáciles para alumnos con NSE Alto y una docena de ítems (SMA\_07, SMB\_08, SMD\_05, SMD\_06, SMD\_09, SME\_06, SMF\_06, SMF\_07, SMF\_10, SMG\_06, SMH\_09 y SMH\_11) con DIF parcial favorable hacia los alumnos con NSE Bajo. Se halló que la mayoría de ítems con DIF parcial ocurren con los alumnos ubicados en el tercil superior de habilidad matemática (ver Anexo 18).



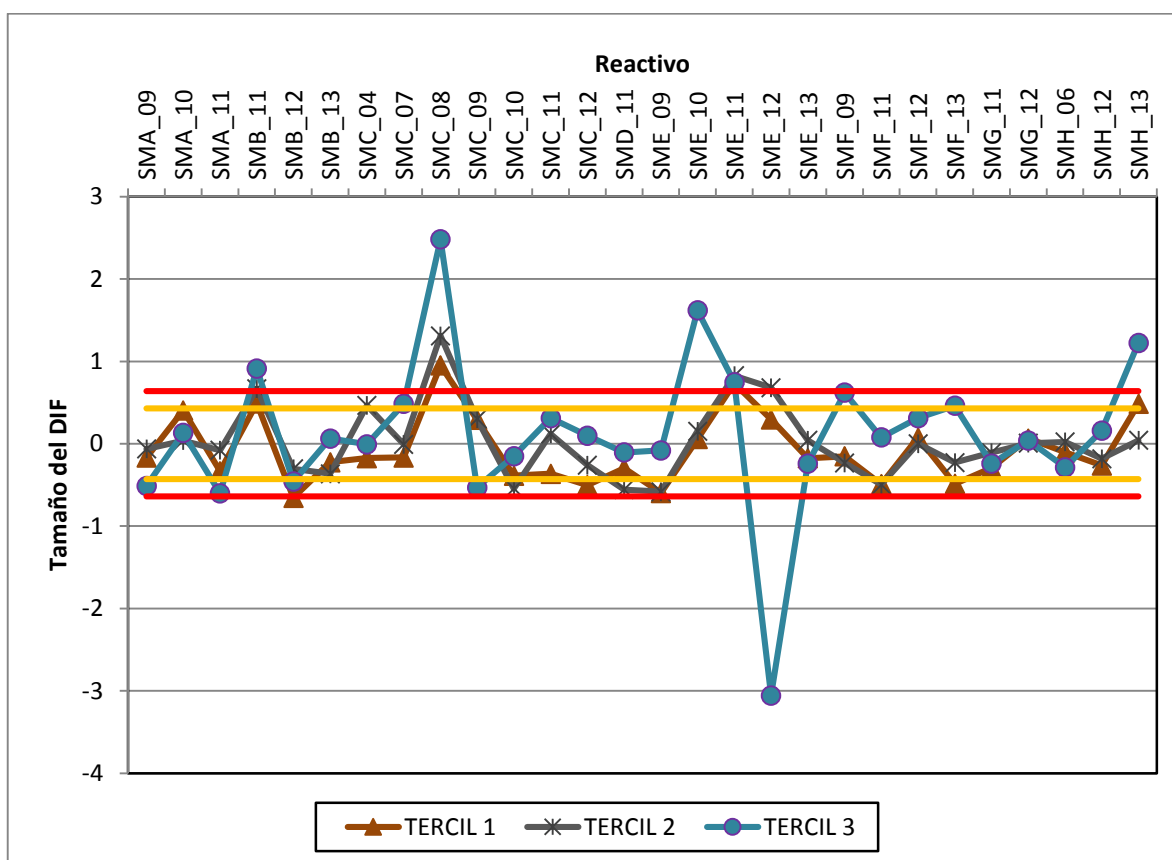
**Figura 13. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Forma, Espacio y Medida* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico**

En la subescala MI, compuesta por 28 ítems, no se detectó ningún ítem con NUDIF severo (ver Figura 14). Sin embargo, al realizar análisis DIF parciales, se encontró que entre la tercera parte de estudiantes con los rendimientos más bajos en Matemáticas (el Tercil 1) hay dos ítems (SMC\_08 y SME\_11) que tienen DIF parcial a favor de los estudiantes con NSE Alto y un ítem (SMB\_12) que resultó más fácil para los alumnos con NSE Bajo.

En el segundo tercil se encontraron cuatro ítems (SMB\_11, SMC\_08, SME\_11 y SME\_12) que tienen un funcionamiento diferencial parcial y todos a favor de



alumnos con NSE Alto. Entre los alumnos que evidenciaron los niveles más altos en la habilidad matemática (tercil superior) se diagnosticaron seis ítems con DIF parciales severos; cinco con DIF parcial favorable hacia estudiantes con NSE Alto (SMB\_11, SMC\_08, SME\_10, SME\_11 y SMH\_13) y el ítem SME\_12 que resultó más fácil para los alumnos provenientes de familias con NSE Bajo (ver Anexo 19).



**Figura 14. Tamaño del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala *Manejo de la Información* del Excale 09 de Matemáticas en 2012, desglosados por Tercil de habilidad y por Nivel Socioeconómico**

#### 4.4. Ítems con sospecha de Sesgo

A partir de los análisis puntuales de DIF que se desarrollaron para el Excale 09 y de cada una de las subescalas de Matemáticas, se identificaron 18 ítems en los que se sospecha que existe sesgo, con base en alguno de los siguientes criterios:

1. Que el ítem presentase DIF Uniforme severo
2. Que el ítem presentase NUDIF severo; es decir, donde en uno de los terciles se presenta DIF enorme a favor de una de las subpoblaciones y en otro de los terciles de habilidad se exhibe DIF severo para una subpoblación distinta a la anterior.

Aplicando estos criterios, los análisis DIF y NUDIF realizados con los cien ítems permiten concluir que ninguno de ellos presentó DIF o NUDIF por sexo, en tanto que se identificaron 18 ítems con sospecha de sesgo entre alumnos de distinto NSE.

La Tabla 16 presenta los 18 ítems con sospecha de sesgo. De la subescala SNPA se identificaron cuatro ítems con DIF severo (SMB\_01, SMB\_03, SME\_02, SMF\_02) y dos con NUDIF severo (SMA\_05, SMG\_05). En la subescala FEM se hallaron cinco ítems con DIF severo (SMB\_09, SMD\_04, SMD\_05, SMG\_07 y SMH\_11) y ninguno con NUDIF. De la subescala MI se encontraron siete ítems con DIF severo (SMB\_11, SMC\_08, SMC\_10, SMD\_11, SME\_09, SME\_11 y SME\_12) y ninguno con NUDIF.

La primera columna de la Tabla 16 presenta las tres subescalas de Matemáticas y la segunda columna los ítems sospechosos de sesgo. En la columna DIF y las tres de NUDIF se señala, con los símbolos que se explican al pie de la tabla, el nivel de sesgo moderado o severo, y su sentido favorable a una u otra de las subpoblaciones consideradas. Nótese que el DIF no es predominante para cierto subgrupo poblacional, sino que aproximadamente la mitad de los ítems favorecen a los alumnos de NSE Alto y la otra mitad a los estudiantes de NSE Bajo. El dictamen de cada ítem se puede ver en el Anexo 20.

**Tabla 16. Resumen de ítems con DIF y NUDIF por Nivel Socioeconómico del Excale 09 de Matemáticas en 2012**

Subescala	Ítem	DIF	NUDIF		
			Tercil 1	Tercil 2	Tercil 3
SNPA	SMA_05		--		++
	SMB_01	++		++	++
	SMB_03	--		-	--
	SME_02	--	-	--	
	SMF_02	++	++	++	++
	SMG_05		++		--
FEM	SMB_09	++	+	++	
	SMD_04	--		--	
	SMD_05	--		--	--
	SMG_07	++		++	++
	SMH_11	--	--	--	--
MI	SMB_11	++	+	++	++
	SMC_08	++	++	++	++
	SMC_10	--		-	
	SMD_11	--		-	
	SME_09	--	-	-	
	SME_11	++	++	++	++
	SME_12	++		++	--

**Fuente:** elaboración propia

**Notas:** + DIF moderado a favor de NSE Alto; ++ DIF severo a favor de NSE Alto  
 - DIF moderado a favor de NSE Bajo; -- DIF severo a favor de NSE Bajo

## **5. RESULTADOS DEL ANÁLISIS DE SESGO MEDIANTE PROTOCOLOS**

### **VERBALES**

En este capítulo se presentan los resultados del análisis de sesgo efectuado con los 18 reactivos que fueron detectados con DIF durante la primera etapa del estudio, y que se describieron en el capítulo 4 de esta tesis. El análisis de sesgo de cada reactivo se localiza en los anexos 20-37 y es un material que puede ser de utilidad para los diseñadores de pruebas, especialmente para los involucrados en el diseño de las pruebas Planea de secundaria para Matemáticas. Este capítulo está organizado en cuatro subcapítulos, en cada uno de ellos se presentan los resultados por cada uno de los cuatro tipos de problemas de sesgo que fueron examinados en esta investigación doctoral, basados en el Diseño Universal de las Evaluaciones (DUE): 1) definición imprecisa del constructo; 2) ítem no accesible o sesgado; 3) procedimientos e instrucciones complejas, poco claras e intuitivas; y 4) incomprensibilidad o problemas de legibilidad del ítem.

En términos generales, de los 18 ítems que se analizaron mediante los protocolos y reportes verbales (Ercikan *et al.*, 2010; Ericsson y Simon, 1993; Fleck y Weisberg, 2004) en 13 de ellos se encontró evidencia de al menos un tipo de sesgo. En la Tabla 17 se presenta la información desglosada por cada reactivo. Podrá notarse, además, que siete ítems se diagnosticaron con incomprensibilidad o problemas legibilidad, cuatro reactivos se detectaron con una definición imprecisa del constructo, la evidencia mostró que seis ítems tienen problemas de accesibilidad para ciertos grupos de informantes y ninguno de los 18 reactivos

examinados tuvo procedimientos e instrucciones complejas, poco claras e intuitivas.

**Tabla 17. Número de reactivos que presentan problemas de sesgo y cantidad de informantes que aportaron evidencias de sesgo, desagregado por tipo**

Folio Reactivo	Respuesta correcta	Tipos de sesgo				Reactivo presenta problema de sesgo
		Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem	
SMA-05	6	0	0	0	0	No
SMB-01	7	0	4	0	3	Sí
SMB-03	7	3	0	0	0	Sí
SMB-09	7	0	0	0	2	Sí
SMB-11	8	0	0	0	0	No
SMC-08	8	0	2	0	0	Sí
SMC-10	4	8	3	0	4	Sí
SMD-04	3	0	7	0	0	Sí
SMD-05	7	0	0	0	0	No
SMD-11	5	0	3	0	1	Sí
SME-02	7	3	0	0	0	Sí
SME-09	1	0	0	0	4	Sí
SME-11	6	0	0	0	2	Sí
SME-12	8	2	0	0	0	Sí
SMF-02	8	0	0	0	0	No
SMG-05	6	0	0	0	0	No
SMG-07	5	0	1	0	0	Sí
SMH-11	0	0	0	0	7	Sí
Núm. de reactivos con presencia de varianza irrelevante al constructo		4	6	0	7	13

Fuente: elaboración propia

### 5.1. Definición imprecisa del constructo

Desde la perspectiva del DUE los constructos que se pretenden medir deberían estar definidos de forma precisa de tal manera que las especificaciones de los ítems y las tareas evaluativas que se derivan de ellas contribuyan en conjunto a medir lo que se desea evaluar. Por ello, desde un enfoque del DUE las evaluaciones debería remover todas aquellas barreras cognitivas, sensoriales, emocionales y físicas que no están relacionadas con el constructo que se desea medir, lo que vuelve imprecisa su medición (Thompson *et al.*, 2002).

A partir de los protocolos y reportes verbales se identificaron cuatro ítems (SMB-03, SMC-10, SME-02 y SME-12) que presentaron una definición imprecisa del constructo; en cada uno de esos reactivos de dos a ocho alumnos y docentes entrevistados mostraron evidencias que apuntan en ese sentido. Además, vale señalar que los participantes que aportaron información sobre la falta de precisión en la medición de estos ítems, provienen de los tres niveles socioeconómicos (NSE) en que se les clasificó, sin que haya diferencias importantes según el tipo de DIF con el que fue diagnosticado el ítem (ver Tabla 18).

**Tabla 18. Número de entrevistados por ítem que aportaron evidencias sobre una definición imprecisa del constructo, por nivel socioeconómico**

Folio Reactivo	Tipo de DIF con el que fue diagnosticado	Nivel Socioeconómico			Total
		Bajo	Medio	Alto	
SME-12	++	1	0	1	2
SMC-10	--	3	3	2	8
SMB-03	--	1	1	1	3
SME-02	--	2	0	1	3
Total		7	4	5	16

**Notas:**

++ DIF severo a favor de NSE Alto

-- DIF severo a favor de NSE Bajo

**Fuente:** elaboración propia

En cada uno de los siguientes dos subapartados se presentan las evidencias más ilustrativas del problema de imprecisión en la medición del constructo, tomando como ejemplo un par de los ítems evaluados.

#### **5.1.1. Imprecisión en la medición del constructo del ítem SMB-03**

El ítem SMB-03 (ver Figura 15) pertenece al eje de la asignatura de Matemáticas llamado *sentido numérico y pensamiento algebraico* (SNPA); en el programa de Matemáticas se ubica dentro del tema *significado y uso de las operaciones* y en el subtema *potenciación y radicación*. El reactivo pertenece a la especificación del Examen para la Calidad y Logro Educativos (Excale) que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen el uso de leyes de exponentes.

Diana resolvió la siguiente expresión:

$$\frac{(4^2)(4)(4^4)(4^0)}{(4^3)(4^5)} = \frac{1}{4^4}$$

La maestra de Matemáticas dijo a Diana que el resultado que obtuvo era incorrecto. ¿Cuál es la solución que Diana debía obtener?

- 0
- 4
- 1
- $\frac{1}{4}$  \*

\* Opción correcta

**Figura 15. Reactivo SMB-03**

De acuerdo con su especificación, para contestar el reactivo de forma correcta se requiere que el sustentante involucre las siguientes actividades cognitivas:

- Identificar que el problema o la expresión que se presenta se resuelven mediante la aplicación de las leyes de los exponentes.
- Aplicar el (los) algoritmo(s) correspondiente(s) de la(s) ley(es) de los exponentes.
- Encontrar la respuesta posible del problema (INEE, 2010c, p. 8).

Además, se necesita que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Resolución de operaciones básicas con números enteros, decimales y fracciones.
- Potenciar números naturales y decimales.
- Razonamiento lógico matemático.
- Interpretar problemas verbales o numéricos.

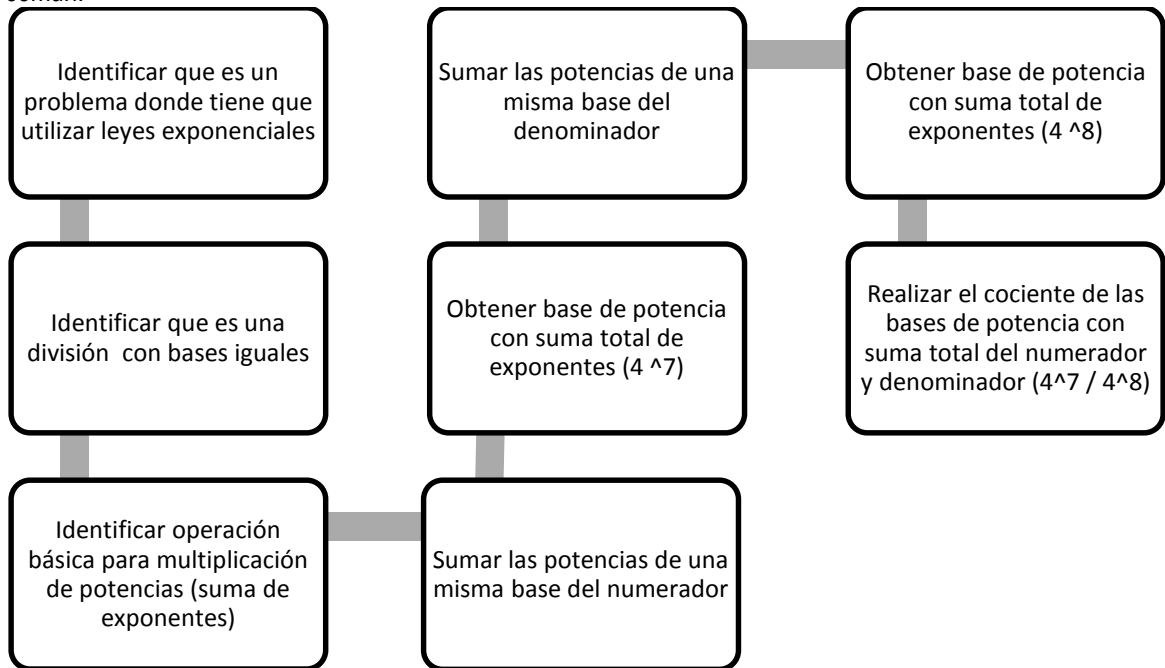


- Aplicación de leyes de los exponentes (INEE, 2010c, p. 8).

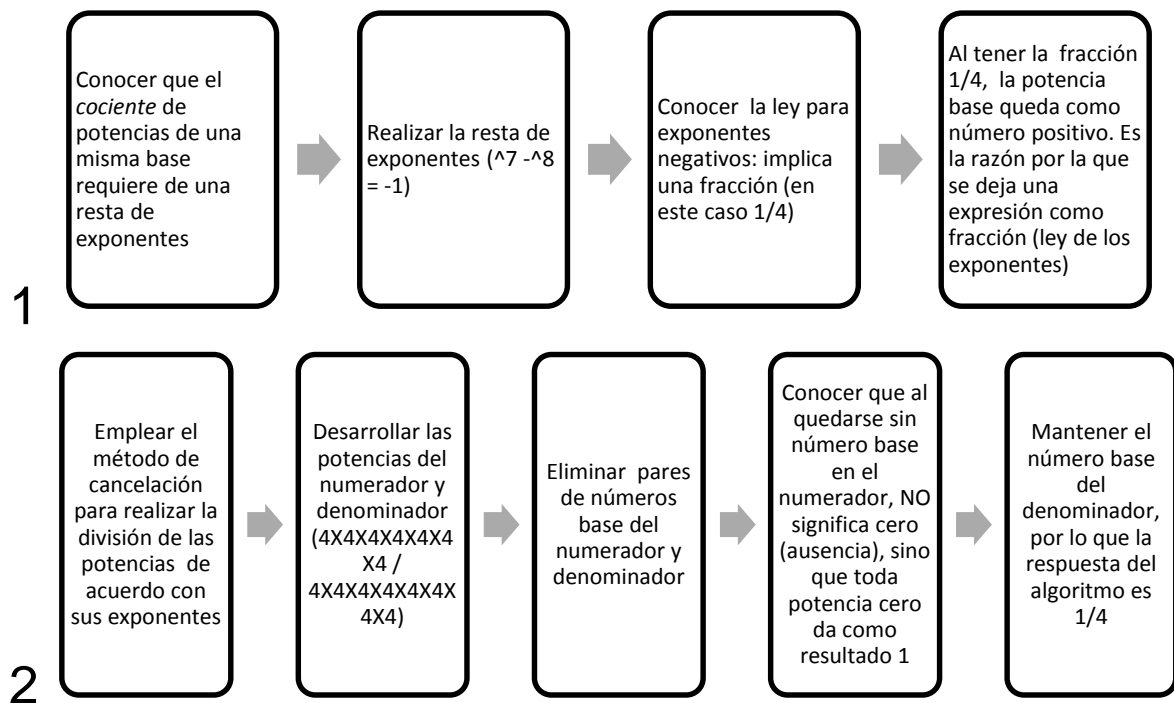
Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes emplearían el siguiente proceso cognitivo, que en un inicio tiene una parte común y posteriormente los entrevistados podrían optar por alguno de los dos procesos señalados a partir de la bifurcación (ver Figura 16).

Un primer análisis para identificar el potencial problema de definición imprecisa del constructo consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían ser capaces de identificar que el problema implica la utilización de las leyes de exponentes, sumar los exponentes al multiplicar potencias de una misma base, restar los exponentes al dividir potencias de una misma base, eso coincide con lo que el ítem pretende medir, si el sustentante es capaz de: potenciar números naturales y decimales; y aplicar leyes de los exponentes.

Parte común:



A partir de la bifurcación:



**Figura 16. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-03**

Un segundo análisis, a partir de los protocolos verbales, permitió identificar tres casos, uno de cada NSE, que muestran evidencia de que el reactivo SMB-03 evoca conocimientos y habilidades distintas a las que pretende medir. En la especificación se establece que este ítem pretende medir si los sustentantes conocen y son capaces de aplicar las leyes de los exponentes; no obstante, tres alumnos que desconocían las leyes de los exponentes siguieron un procedimiento distinto mediante el cual se puede llegar la respuesta correcta.

Uno de los alumnos (de NSE alto) que no conocía las leyes de los exponentes y aun así seleccionó la opción correcta, cuando se le pregunta si comprendió el problema responde:

sí, pero no sé qué tengo que hacer... no sé si éstos [refiriéndose a los términos con exponentes] los tengo que multiplicar... la verdad no sé qué hacer (E08).

Pese a ello, este alumno pudo resolver correctamente el reactivo siguiendo un método que implica convertir los exponentes en números naturales (ver Figura 17). Este entrevistado llegó al resultado correcto sin aplicar alguna de las leyes de los exponentes, que es lo que el ítem pretende medir; en cambio, fue convirtiendo (incluso con errores que luego se compensaron) los números con notación potencial a números naturales, desarrollando y resolviendo la siguiente expresión:

$$\frac{(4^2)(4)(4^4)(4^0)}{(4^3)(4^5)} = \frac{(8)(4)(128)(1)}{(32)(512)} = \frac{(8)(512)}{16384} = \frac{4096}{16384} = 0.25 = \frac{1}{4}$$

La maestra de Matemáticas dijo a Diana que el resultado que obtuvo era incorrecto. ¿Cuál es la solución que Diana debía obtener?

0  
 -4  
 1  
  $\frac{1}{4}$

The image shows handwritten mathematical work. At the top left, there are two expressions:  $(8)(4)(128)$  and  $(32)(512)$ . Below these are several calculations. On the left, there is a multiplication:  $128 \times 4 = 512$ , followed by  $1024$  and  $1536$ . In the center, there is a multiplication:  $16384$  and  $32768$ . On the right, there is a division:  $4096 \div 16384 = 0.25$ . The student has also written  $16,384$  and  $71,920$  in various places.

Figura 17. Operaciones realizadas por E08 para responder ítem SM-03

Nótese que el alumno obtiene un resultado correcto a pesar de haber cometido errores en la conversión de potencias a números naturales porque un error se compensa con otro. El alumno al final divide 4096 (resultado de multiplicar  $(8)(4)$   $(128)(1)$ ) entre 16384 (resultado de multiplicar  $(32)(512)$ ) y obtiene 0.25. Obviamente  $4^2$  no es 8, sino 16;  $4^4$  es 256 y no 128;  $4^3$  es 64 y no 32, y  $4^5$  es 1024 y no 512. Lo correcto sería poner en el numerador el producto de  $(16)(4)(256)(1)$

y en el denominador (64) (1024). El resultado es la fracción  $16384/65536$  que también resulta 0.25.

Los otros dos entrevistados que también aportaron evidencias sobre una definición imprecisa del constructo, pese a que uno de ellos no respondió correctamente el reactivo y el otro al final adivinó y acertó la respuesta correcta, siguieron el mismo procedimiento que el alumno anterior; sin embargo, al desarrollar las expresiones cometieron varios errores aritméticos. Nótese en la Figura 18 cómo otro de los alumnos (E05) utilizando el mismo procedimiento de convertir las potencias a notación natural, a pesar de tener una comprensión errónea de lo que significa elevar un número a la  $n$  potencia (pues al tratar de elevar  $4^5$  lo realiza de la siguiente manera  $4 \times 4 = 16$ , luego opera  $16 \times 4 = 64$ , después comete el error cuando multiplica  $64 \times 64$  [lo correcto sería  $16 \times 4$ ] y posteriormente multiplica  $4296 \times 4$ ), y a pesar de cometer errores aritméticos (como cuando al multiplicar  $64 \times 64$  obtiene por resultado 4296 cuando en realidad es 4096) selecciona la respuesta correcta, pues utilizó las opciones del ítem para comparar sus resultados parciales, de tal manera que fue ajustando su resultado, hasta que encontró una opción que lo satisfizo en función de las operaciones que fue empleando y de lo que el problema planteaba resolver.

El otro alumno que utilizó este mismo procedimiento cometió errores aritméticos, de haber realizado correctamente las operaciones aritméticas para convertir de números exponenciales a números naturales, muy posiblemente lo haya respondido de manera correcta, lo cual implica que el ítem puede ser resuelto empleando habilidades y evocando conocimientos diferentes a los planteados por

el constructo que se desea medir, que en este caso son las leyes de los exponentes. En otras palabras, estos tres entrevistados siguieron un procedimiento de resolución que no corresponde ni con las habilidades establecidas en la especificación del ítem ni con los procesos cognitivos que anticiparon las especialistas, a diferencia de los demás entrevistados que lo respondieron de manera correcta.

La maestra de Matemáticas dijo a Diana que el resultado que obtuvo era incorrecto. ¿Cuál es la solución que Diana debía obtener?

0  
 -4  
 1  
  $\frac{1}{4}$

$$\frac{(12)(4)(64)(4)}{(64)(1024)}$$

$$\begin{array}{r} 64 \\ \times 4 \\ \hline 256 \\ \times 4 \\ \hline 1024 \end{array}$$

$$\begin{array}{r} 16 \\ \times 4 \\ \hline 64 \\ \times 64 \\ \hline 256 \\ \times 4 \\ \hline 404 \\ \times 4 \\ \hline 17184 \end{array}$$

$$\begin{array}{r} 1024 \\ \times 64 \\ \hline 4096 \\ 6044 \\ \hline 64536 \end{array}$$

$$\frac{1024^7}{16^8} =$$

$$\begin{array}{r} 69 \\ 16 \overline{)1024} \\ \underline{96} \\ 164 \\ \underline{144} \\ 020 \end{array}$$

Figura 18. Operaciones realizadas por E05 para responder ítem SM-03

### 5.1.2. Imprecisión en la medición del constructo del ítem SMC-10

El ítem SMC-10 (ver figura 19) pertenece al eje de la asignatura de Matemáticas llamado *manejo de la información* (MI); en el programa de Matemáticas corresponde al tema *representación de la información* y al subtema *gráficas*. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen la interpretación de información representada en gráficas de línea.

De acuerdo con su especificación, para resolver el ítem se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- En primer lugar, los alumnos deben comprender la información contenida en la gráfica. La comprensión de la gráfica es un proceso complejo que implica identificar claramente los elementos estructurales que, según Friel, Curcio y Bright (2001), son:
  - Contenido contextual: saber leer el título y etiquetas (para comprender qué información se muestra en la gráfica y cuáles son las variables representadas)
  - Marco del gráfico: saber identificar los ejes interpretar correctamente la escala usada en cada uno (para comprender las unidades de medida de las magnitudes representadas en el gráfico). Por lo que la interpretación de los números (naturales, fracciones comunes decimales) en la recta numérica es un conocimiento esencial para dar respuesta al reactivo.
  - Especificadores del gráfico: saben interpretar que, en el caso de una gráfica de línea, las coordenadas de los puntos unidos por los segmentos que forman la línea, son parejas ordenadas cuyo primer elemento corresponde a la unidad del tiempo y el segundo elemento corresponde a la magnitud graficada y representada en el eje

vertical. Esto implica que los alumnos estén familiarizados con la representación de puntos en un plano cartesiano.

- Para el caso del contenido que se evaluará y como se presentará una gráfica de línea con dos magnitudes que varían con respecto al tiempo, para responder al reactivo se tendrán que reconocer los elementos estructurales mencionados en ambas magnitudes por separado y, además, relacionándolos entre sí.
- Si se decide usar porcentajes como unidad del eje vertical, el manejo de los mismos constituye un conocimiento necesario para dar respuesta al reactivo (INEE, 2010g, p. 6).

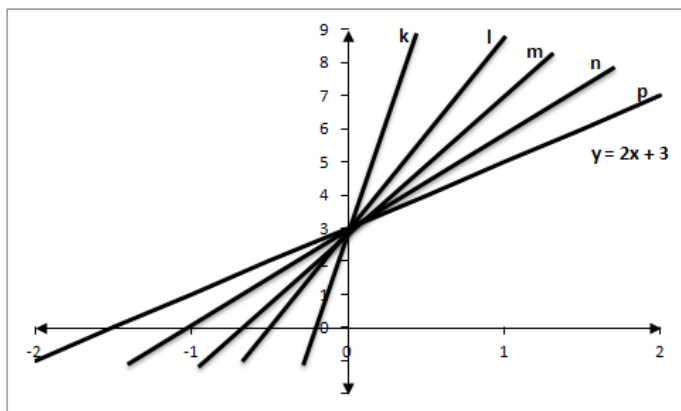
Además, las actividades cognoscitivas que se prevén en la especificación que estén involucradas en la solución correcta del reactivo son:

- Interpretar la información contenida en una gráfica de línea (Título, etiquetas de los ejes, escalas, etc.).
- Identificar la información necesaria para dar respuesta a la pregunta planteada.
- Analizar la información identificada(INEE, 2010g, p. 6).



Considera las siguientes gráficas de la forma  $f(x) = mx + b$ .

La ecuación de la recta p es  $y = 2x + 3$ .



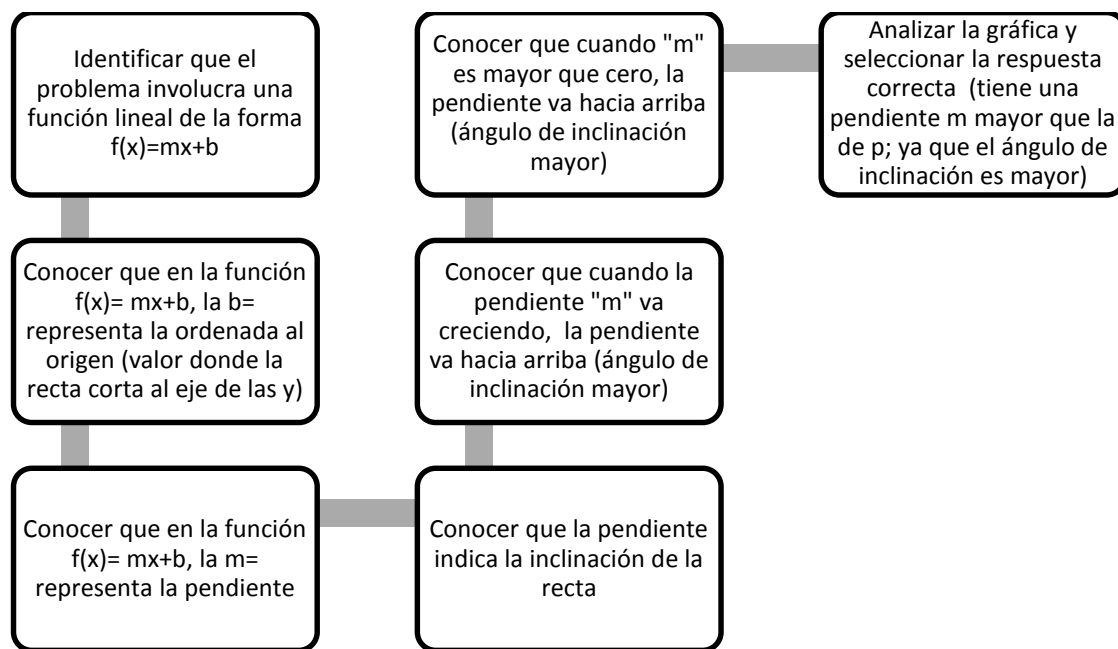
¿Cuál de las siguientes afirmaciones describe lo que ocurre con las rectas k, l, m, n con respecto de la recta p?

- Tienen una pendiente m igual que la de p, ya que pasan por un mismo punto.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es mayor.\*
- Tienen una pendiente m menor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es menor.

\* Opción correcta

**Figura 19. Reactivo SMC-10**

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un proceso cognitivo lineal como el que se ilustra en la Figura 20.



**Figura 20. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMC-10**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas se encontró inconsistencia entre ambas fuentes. Mientras que las especialistas anticiparon un proceso que implicaba el conocimiento de la pendiente ( $m$ ) y la ordenada al origen ( $b$ ) en la función  $f(x)=mx+b$  así como conocer el comportamiento gráfico de la pendiente cuando el ángulo de inclinación es mayor que cero; en cambio, en la especificación del reactivo si bien se señala que los alumnos deben comprender la información contenida en la gráfica, especifican que los alumnos deben: interpretar correctamente la escala usada en cada uno de los ejes del plano cartesiano; interpretar los números (naturales, fracciones comunes decimales) en la recta numérica; interpretar que, en el caso de una gráfica lineal, las coordenadas son parejas ordenadas, cuyo primer elemento corresponde a la

unidad del tiempo y el segundo elemento corresponde a la magnitud graficada y representada en el eje vertical, entre otras habilidades.

Un segundo análisis, a partir de la información recabada mediante los protocolos verbales, permitió conocer que para enfrentarse a la tarea evaluativa ocho de los entrevistados emplearon un proceso cognitivo que es distinto con los conocimientos y habilidades que se plantean en la especificación del reactivo, es decir, la evidencia apunta a una definición imprecisa del constructo que se desea evaluar.

Si bien los alumnos no respondieron de forma correcta el ítem, cuando se enfrentaron a él, intentaban evocar conocimientos como pendiente, ordenada al origen, ecuaciones de la forma  $f(x) = mx + b$  y ángulo de inclinación. Los docentes quienes respondieron el reactivo de forma correcta, evocaron estas cuestiones de manera adecuada.

En los protocolos, se identificaron muchos episodios que ilustran cómo los alumnos y docentes utilizan conocimientos y habilidades distintos a los señalados en la especificación del reactivo, para tratar de encontrar la respuesta correcta. Por ejemplo, cuando una de las alumnas al ir analizando las opciones de respuesta y descartar la opción b) (que es la opción correcta) afirma que esa opción no puede ser la correcta porque "...tienen una pendiente  $m$  mayor que la de  $p$ ..." (E02).

O bien, cuando a otra de las alumnas se le pide que explique por qué seleccionó la opción a) que incorrecta, justifica:

porque tienen una pendiente igual, porque aquí se observa [mientras compara la longitud de las rectas  $p$  y  $m$ ], y si las juntamos, es igual, nada más que una está más inclinada que la otra... y las dos pasan por un mismo punto... (E04).

Un elemento fundamental para responder correctamente el reactivo es el concepto de pendiente, conocimiento distinto a lo que plantea la especificación a partir del cual se diseñó el reactivo. El entrevistado que aparentemente adivinó manifiesta su confusión para responder al ítem de la siguiente manera:

Bueno, la palabra que no entiendo es ¡pendiente!... tiene una pendiente  $m$ , eso no [...] ...es que no sé qué es la pendiente... (E05).

Por su parte, los docentes, que respondieron de manera correcta el ítem, también evocaron conocimientos distintos a los previstos en la especificación del reactivo. Se ilustra cuando uno de ellos verbaliza:

La dificultad de este reactivo estriba en recordar el concepto de pendiente... muchas veces son conceptos teóricos que a la hora de verlos en un esquema, en un diagrama, sí se les dificulta recordar qué es la pendiente (E03).

## **5.2. Ítem no accesible o sesgado**

En el campo de las evaluaciones en gran escala, la accesibilidad se comprende como la oportunidad y sin obstáculos que “debe ofrecer una prueba a todos los sustentantes para demostrar el estado que tienen con respecto al constructo que pretende medir” (AERA-APA-NCME, 2014: 49). Un ítem se considera sesgado

cuando incluye contenido que podría resultar ofensivo o penalizar a ciertos grupos de estudiantes debido a sus características personales como raza, género, etnicidad, discapacidad o estatus socioeconómico, los cuales representan aspectos distintos a lo que se pretende medir y, por tanto, no debieran favorecer u obstaculizar a los sustentantes en función de estas variables. Se parte del supuesto de que algunos rasgos de una prueba pueden dificultar que algunos sustentantes muestren el manejo que tienen de lo que se quiere medir, por ejemplo, cuando el diseño tipográfico no permite que los sustentantes con alguna debilidad visual puedan leer los ítems de una prueba que no está destinada a medir la capacidad visual (Thompson *et al.*, 2002).

A partir de los protocolos verbales se identificaron seis ítems que presentaron problemas de inaccesibilidad o sesgo. Tres de ellos favorecen a alumnos de NSE alto (SMB-01, SMC-08 y SMG-07) y la otra mitad a alumnos de NSE bajo (SMD-04, SMC-10 y SMD-11); en cada uno de esos reactivos, de uno a siete alumnos y docentes entrevistados mostraron evidencias que apuntan en ese sentido. Los participantes que aportaron información sobre la falta de precisión en la medición de estos ítems, provienen de los tres NSE en que se les clasificó, aunque más de la mitad (55%) de las evidencias provienen de entrevistados de NSE bajo, 35% de ellas provienen de participantes de NSE medio y sólo 10% de las evidencias de alumnos y docentes de NSE alto, lo cual de antemano denota que si bien los reactivos con este tipo de sesgo favorecen a alumnos de NSE alto y bajo, se encontraron más problemas de inaccesibilidad para la población de NSE bajo (ver Tabla 19).

**Tabla 19. Número de entrevistados por ítem que aportaron evidencias sobre inaccesibilidad o sesgo, por nivel socioeconómico**

Folio Reactivo	Tipo de DIF con el que fue diagnosticado	Nivel Socioeconómico			Total
		Bajo	Medio	Alto	
SMB-01	++	2	2	0	4
SMC-08	++	0	1	1	2
SMG-07	++	1	0	0	1
SMD-04	--	3	3	1	7
SMC-10	--	3	0	0	3
SMD-11	--	2	1	0	3
Total		11	7	2	20

**Notas:**

++ DIF severo a favor de NSE Alto

-- DIF severo a favor de NSE Bajo

**Fuente:** elaboración propia

Para ilustrar el problema de inaccesibilidad o sesgo en los reactivos, en los siguientes dos subapartados se presentan las evidencias más ilustrativas tomando como ejemplo un par de los ítems evaluados.

### 5.2.1. Inaccesibilidad o sesgo del reactivo SMB-01

El ítem SMB-01 (ver Figura 21) pertenece al eje de la asignatura de Matemáticas llamado *sentido numérico y pensamiento algebraico*; en el programa de Matemáticas se ubica dentro del tema *significado y uso de las operaciones* y el subtema *problemas aditivos*. En el Excale, este reactivo pertenece a la especificación denominada resolver problemas aditivos que impliquen el uso de números con signo (INEE, 2010b).

En el poblado donde vivo se registró una temperatura de  $12^{\circ}\text{C}$  el domingo a mediodía; durante las siguientes 18 horas la temperatura descendió  $15^{\circ}\text{C}$ . ¿Qué temperatura marcó el termómetro después de las 18 horas transcurridas?

- $-3^{\circ}\text{C}$  \*
- $3^{\circ}\text{C}$
- $27^{\circ}\text{C}$
- $-27^{\circ}\text{C}$

\* Opción correcta

**Figura 21. Reactivo SMB-01**

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Sumar y restar números naturales, fraccionarios y decimales positivos o negativos.
- Conocer la secuencia numérica ascendente y descendente a partir de cualquier número.
- Utilizar criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma o con una resta.
- Calcular sumas y restas apoyándose en la recta numérica.
- Resolver problemas aditivos de cambio y de composición (INEE, 2010b, p. 6).

Cuatro de los entrevistados (3 que lo respondieron correctamente y 1 que no seleccionó la respuesta correcta), aportaron elementos que apuntan que el

reactivo SMB-01 no guarda uno de los principios fundamentales del DUE: la accesibilidad e imparcialidad.

Para el grupo de NSE bajo, la utilización de ciertos términos en el pie del ítem puede obstaculizar que los sustentantes muestren sus habilidades matemáticas a cabalidad. El proceso empleado por uno de los estudiantes de TVS es muy ilustrativo de cómo el desconocimiento de la palabra “descendió” ocasionó que se eligiera una opción incorrecta. Este alumno, después de leer el pie del ítem, reconoce que se trata de un problema de adición y resuelve la operación  $12+15=27$ , luego selecciona la opción correspondiente a  $27^{\circ}\text{C}$ , que es errónea.

Este alumno no consideró de manera correcta el signo de las temperaturas, porque para el entrevistado la palabra “descendió” significó un aumento en la temperatura; cuando verbalizaba su comprensión de la tarea evaluativa, dijo: “Aquí observo que aumentó tres grados durante esas 18 horas” (E01).

De hecho, el entrevistado señala que como la temperatura inicial es de  $12^{\circ}\text{C}$  y luego aumenta  $15^{\circ}\text{C}$ , las opciones de respuesta con números negativos no pueden ser las correctas; esto se ilustra cuando dijo: “...menos tres grados centígrados no puede ser, si la temperatura está aumentando...” (E01).

Su comprensión errónea de la palabra “descendió” se pudo corroborar en otras partes del reporte verbal, por ejemplo, cuando explica cómo descartó algunas de las opciones de respuesta señala que si el reactivo plantease “un incremento proporcional de  $3^{\circ}\text{C}$  por cada hora”, algunos alumnos también podrían elegir la



opción de 3°C, sin embargo, el entrevistado señala que eso no es lo que plantea el problema, y por lo cual no seleccionó la segunda opción.

La evidencia obtenida a través de la técnica de pensamiento en voz alta, también permitió conocer que el estudiante disponía de los conocimientos previos señalados en la especificación del reactivo, tales como suma de números naturales, utilizaba criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma, y resolución de problemas aditivos de cambio y de composición; sin embargo, la inclusión en el reactivo de una palabra que es poco familiar para el estudiante, y la cual es clave para la comprensión cabal del problema a resolver, tal como lo señaló una de las docentes, y del procedimiento matemático que debe realizar, se convirtió en un obstáculo para que el alumno escogiera la opción correcta.

El desconocimiento o comprensión de una palabra, tal como “descendió”, para ciertos grupos de población, denota que el reactivo si bien mide aspectos sobre el conocimiento matemático, también mezcla cuestiones de vocabulario, que corresponde más bien a otro de los subdominios que evalúa el Excale 09 para la asignatura de Español: comprensión lectora o eventualmente reflexión sobre la lengua. Y esta situación no es deseable.

Por otro lado, el reactivo al parecer también puede ser respondido correctamente a partir de la información del contexto de ciertos subgrupos poblacionales. Una de las alumnas (de NSE medio) que lo respondió de forma correcta, pero que no utilizó algún procedimiento matemático para llegar a la solución, utilizó su

conocimiento empírico sobre las temperaturas durante el día y la noche, para acotar las opciones sobre las cuales adivinó la respuesta. Cuando se le preguntó a la estudiante porque seleccionó la opción A (la correcta), respondió:

cambió... al ser de día, va oscureciendo, va oscureciendo y pues va bajando más la temperatura, porque en el día hace más calor y cuando va anocheciendo baja la temperatura y hace más frío (E04).

Si se toma en cuenta que durante su proceso para la resolución del problema no realizó ninguna resta y además, ella dudaba entre las dos opciones que tenían números negativos ( $-3^{\circ}$  y  $-27^{\circ}$ ), se puede inferir que ella dedujo la respuesta correcta, a partir de su conocimiento del medio. La alumna verbalizó “yo creo que es ésta [refiriéndose a las opción a)]... pero también podría ser ésta [señalando la opción d)]” (E04).

### **5.2.2. Inaccesibilidad o sesgo del reactivo SMD-04**

El ítem SMD-04 (ver Figura 22) pertenece al eje de la asignatura de Matemáticas llamado *forma, espacio y medida* (FEM); en el programa de Matemáticas corresponde al tema *medida* y al subtema *estimar, medir y calcular*. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de realizar conversiones de medida de superficie.

Uno de los países más pequeños del mundo es Mónaco, con 2 km<sup>2</sup> de superficie total. ¿A cuántas hectáreas equivale la superficie de este país?

- 20 ha
- 200 ha\*
- 2 000 ha
- 20 000 ha

\* Opción correcta

**Figura 22. Reactivo SMD-04**

De acuerdo con su especificación, para resolver el ítem se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos: “Se debe tener el conocimiento de las unidades de medida de superficies, su equivalencia entre las diferentes unidades de área y la habilidad para realizar conversiones a otras unidades de área” (INEE, 2010e, p. 4).

La evidencia recolectada mediante la técnica de pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMD-04 así como los sondeos que se hicieron durante las entrevistas concurrentes, permitieron encontrar una posible fuente de sesgo en este reactivo.

Siete de los entrevistados (de los tres NSE) tuvieron dificultades o advirtieron que sus alumnos podrían tener dificultades para saber la equivalencia de una hectárea. El conocimiento de la equivalencia es indispensable para resolver correctamente el ítem. En este trabajo cinco de los alumnos y un docente no

supieron dicha equivalencia y ese fue un factor fundamental para que no respondieran correctamente el reactivo.

Al responder un sondeo del entrevistador sobre si le queda clara la pregunta, uno de los alumnos responde “sí, pero no me queda clara la equivalencia para convertirlo a hectáreas” (E08). Otro alumno, después de leer el pie del ítem, trata de recordar cuánto es un kilómetro y cuánto es una hectárea, y dice: “no estoy muy seguro... una hectárea serían... mmm...” y al no responder, la entrevistadora lo sondea cuestionándole si la pregunta le resulta clara, a lo que el mismo alumno responde “sí es clara, sólo que se tiene que recordar lo de kilómetros y de hectáreas” (E05).

Una de las docentes señala que la hectárea no es una medida común en su contexto escolar:

...son medidas que no se manejan mucho. Se maneja el centímetro, se maneja el metro lineal, el kilómetro, pero la hectárea no mucho. Porque una hectárea se maneja más bien, hablando de superficies, de campo, entonces el niño no está muy familiarizado con eso (E06).

Además de la falta de familiaridad en el mundo real con esta medida, salvo para aquéllos que provienen de contextos rurales en los que se manejan las hectáreas para medir superficies de parcelas o ranchos, el docente que atiende alumnos de NSE bajo (y que acuden a una escuela ubicada en un contexto semi-rural) señala que el reactivo será complicado porque este tema suele estar aislado en el programa de matemáticas:

así como está, es un tanto complejo para el alumno porque no... como le digo, es algo que no se retoma mucho a lo largo de los contenidos. Estos contenidos se abordan y ya no se vuelven a aplicar, ese es el problema, que no se aplican en otras secuencias, en otros bloques. Entonces es algo que quedó aislado y que no se le dio aplicación (E03).

Dos de los docentes señalan que si el reactivo incluyese la equivalencia de las hectáreas, por ejemplo, en metros cuadrados, el alumno podría hacer la conversión, porque es más común para ellos saber que 1 km es igual a 1000 metros y a partir de ahí hacer las conversiones necesarias.

El reactivo SMD-04 presentó un DIF que favorece a los alumnos de NSE bajo; a partir de la información de los protocolos y reportes verbales, y tomando en cuenta que existe una fuerte asociación entre NSE y ruralidad, es verosímil razonar que muy probablemente el sesgo del reactivo se derive de aquéllos alumnos que conocen la equivalencia de una hectárea, medida que suele utilizarse en medios rurales y por tanto se pueda inferir que sean los alumnos ajenos a los contextos rurales los que se vieron obstaculizados para responder correctamente el reactivo, a pesar de que podrían haber respondido un ítem similar que le solicitara convertir superficies.

### **5.3. Procedimientos e instrucciones complejas, poco claras e intuitivas**

Desde la perspectiva del DUE, las instrucciones de las pruebas así como los procedimientos sobre cómo responder deberían ser fáciles de entender para los sustentantes, sin importar su experiencia, conocimiento, dominio del lenguaje o

nivel de concentración (Johnstone, 2003; Thompson *et al.*, 2002). Instrucciones muy extensas formuladas con un lenguaje complejo podrían convertirse en una fuente de invalidez porque el sustentante no comprendió cómo debería responder, incluso cuando sí tuviera la habilidad que el test pretendía evaluar.

Otro posible problema en las instrucciones es la variación en el tipo de reactivo, es decir, cuando en una sección del cuadernillo del examen se le pide que rellene el inciso de la respuesta correcta y en la siguiente sección se le solicita que coloque un círculo sobre la opción que seleccione. Por el contrario, cuando en un test de comprensión de lectora se le pide al estudiante que identifique el número de párrafo en el que se presenta la idea principal del texto y los párrafos están numerados, de tal manera que se eliminen las posibilidades de que el sustentante elija un párrafo distinto al que seleccionó como respuesta correcta.

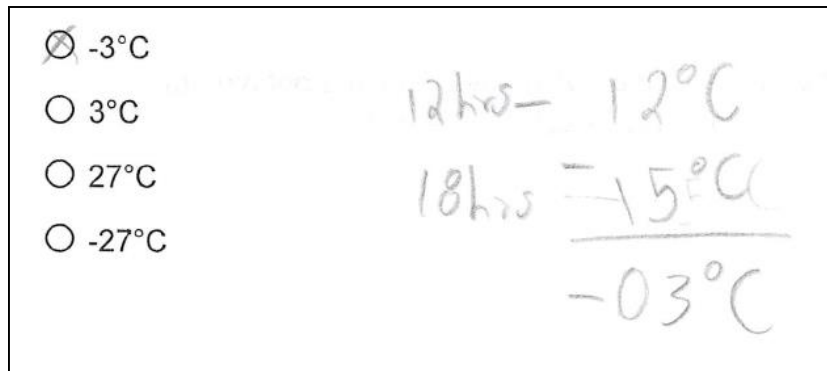
En esta investigación, con ninguno de los 18 ítems se obtuvieron evidencias de que los sustentantes tuvieran dificultades con las instrucciones que les indicaban cómo debían registrar la respuesta a las tareas evaluativas. Más bien, los alumnos y docentes que fueron entrevistados, parecían familiarizados con responder ítems donde deben seleccionar sólo una de las opciones de respuesta que se les ofrece, la cual asumen que es la correcta. Para registrar su respuesta debían rellenar o marcar el alveolo situado al lado izquierdo de cada una de las opciones, como lo ilustra la Figura 23 en la que se recupera una sección de uno de los reactivos en el cuadernillo ya respondido por un estudiante de NSE bajo.

¿Cuál de las siguientes afirmaciones describe lo que ocurre con las rectas k, l, m, n con respecto de la recta p?

- Tienen una pendiente m igual que la de p, ya que pasan por un mismo punto.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m menor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es menor.

**Figura 23. Recorte de un ítem respondido por uno de los alumnos de NSE bajo**

Uno de los entrevistados en lugar de rellenar los alveolos, los tachaba (ver Figura 24). Eso sería un problema en una aplicación de gran escala, donde se utilizan hojas preparadas para lectura óptica o escaneo digital; sin embargo, en el marco de este estudio cognitivo eso no se consideró como evidencia de que las instrucciones fueran complejas, poco claras e intuitivas, puesto que como parte de las adecuaciones que se le hicieron a la prueba para efectos de los protocolos verbales, la primera de las instrucciones señalaba: “Responde cada una de las siguientes preguntas, marcando el óvalo correspondiente para seleccionar la respuesta correcta. Si necesitas hacer operaciones, las puedes realizar en el espacio disponible en la página”; es decir, se le pedía que *marcara* el óvalo correspondiente y no necesariamente que los rellenara. Además, al cuestionarles a los entrevistados la claridad sobre las instrucciones para responder el cuadernillo, todos manifestaron que les resultaron claras.



**Figura 24. Recorte de un ítem respondido por uno de los alumnos de NSE medio**

En resumen, con ninguno de los 18 ítems que se analizaron mediante los protocolos y reportes verbales se encontró evidencia que indicase la presencia de procedimientos e instrucciones complejas, poco claras e intuitivas.

#### **5.4. Incomprensibilidad o problemas de legibilidad del ítem**

Las pruebas diseñadas en el marco del DUE reducen la “complejidad verbal y organizacional de los ítems mientras preservan su contenido esencial (i.e. las habilidades y conceptos que intentan medir) (Hanson *et al.*, 1998, p.2 citado en Thompson *et al.*, 2002). Para ello, una de las estrategias que se utilizan en pruebas que no pretenden medir el dominio de una lengua, como matemáticas o ciencias naturales, es redactar los ítems y las instrucciones de la prueba empleando un lenguaje plano.

Un lenguaje plano es “un lenguaje basado en el texto que es sencillo, conciso, y usa palabras de uso cotidiano para transmitir su significado. La meta de las estrategias de edición del lenguaje plano es mejorar la comprensibilidad del texto



escrito mientras se preserva la esencia del mensaje” (Hanson *et al.*, 1998, p.2 citados en Thompson *et al.*, 2002).

A partir de los protocolos y reportes verbales se identificaron siete ítems que presentaron problemas de incomprensibilidad o legibilidad. Tres de ellos con DIF que favorece a los sustentantes de NSE alto (SMB-01, SMB-09, SME-11 y SME-12) y cuatro reactivos con DIF a favor de alumnos de NSE bajo (SMH-11, SMC-10, SME-09 y SMD-11); en cada uno de esos reactivos, de dos a siete alumnos y docentes entrevistados mostraron evidencias que apuntan en ese sentido. Además, vale señalar que los participantes que aportaron sobre la falta de precisión en la medición de estos ítems, provienen de los tres NSE en que se les clasificó, sin que haya diferencias importantes según el tipo de DIF con el que fue diagnosticado el ítem (ver Tabla 20).

En cada uno de los siguientes dos subapartados se presentan las evidencias más ilustrativas del problema de incomprensibilidad o legibilidad, tomando como ejemplo un par de los ítems examinados.

**Tabla 20. Número de entrevistados por ítem que aportaron evidencias sobre incomprendibilidad o legibilidad del ítem, por nivel socioeconómico**

Folio Reactivo	Tipo de DIF con el que fue diagnosticado	Nivel Socioeconómico			Total
		Bajo	Medio	Alto	
SMB-01	++	1	1	1	3
SMB-09	++	0	1	1	2
SME-11	++	1	1	0	2
SMH-11	--	2	2	3	7
SMC-10	--	0	3	1	4
SME-09	--	2	1	1	4
SMD-11	--	0	1	0	1
Total		6	10	7	23

**Notas:**

++ DIF severo a favor de NSE Alto

-- DIF severo a favor de NSE Bajo

**Fuente:** elaboración propia

#### 5.4.1. Incomprendibilidad o problemas de legibilidad del reactivo SMH-11

El ítem SMH-11 (ver figura 11) pertenece al eje de la asignatura de Matemáticas llamado *forma, espacio y medida*; en el programa de Matemáticas corresponde al tema *formas geométricas* y al subtema *figuras planas*. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de identificar las posibilidades de construcción (existencia y unicidad) de círculos (INEE, 2010d).

Los pueblos de San Agustín y Las Flores se encuentran ubicados a 2 km de distancia entre sí.



¿Cuántas circunferencias de 4 km de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?

- Una
- Dos\*
- Más de dos
- Ninguna

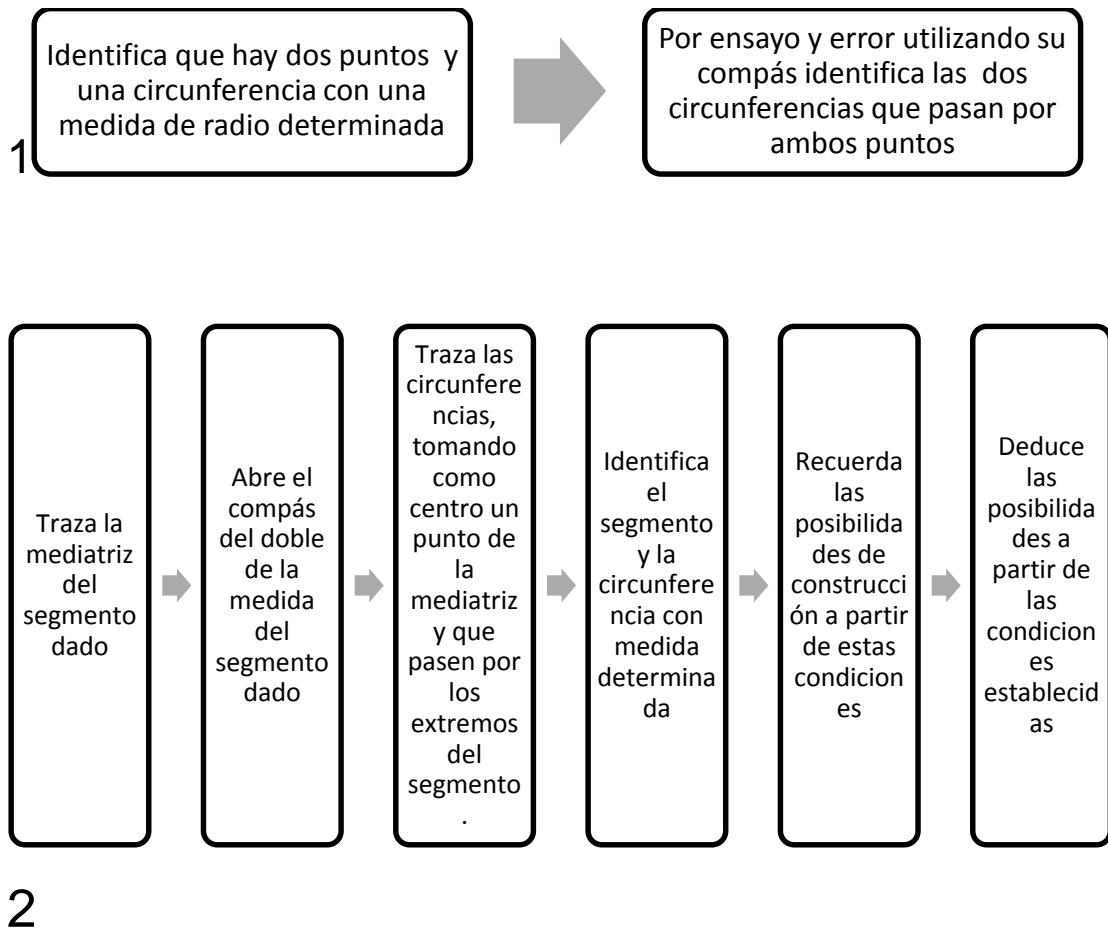
\* Opción correcta

**Figura 25. Reactivo SMH-11**

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante disponga de “los conocimientos sobre las condiciones de posibilidad y unicidad en la construcción de círculos y circunferencias dados algunos datos. Además se debe contar con los conocimientos y habilidades sobre construcción y trazo de figuras geométricas como mediatrices, rectas paralelas y rectas perpendiculares” (INEE, 2010d, pp. 5–6).

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguna de las dos rutas cognitivas que se ilustran en la

Figura 26. Vale decir que tanto en el Excale como en la aplicación de los protocolos verbales los alumnos no utilizan su juego de geometría, por tanto, los entrevistados dibujaron las circunferencias con medidas aproximadas:



**Figura 26. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMH-11**

El análisis de las videograbaciones de los protocolos y reportes verbales aporta elementos para afirmar sobre un posible problema de incomprendibilidad o legibilidad del ítem. Lo que se encontró es que cuatro de los alumnos (de los tres NSE) así como los tres docentes comprendieron erróneamente que el problema

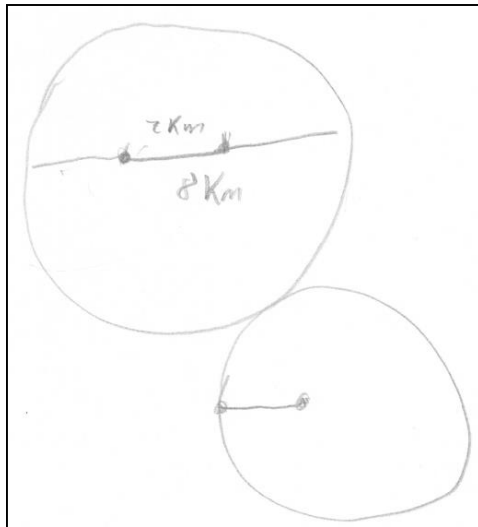
les planteaba encontrar cuántas circunferencias se pueden formar y que cumplieran con las siguientes condiciones: 1) que pasen por San Agustín y Las Flores al mismo tiempo; 2) que tengan 4 km de radio; y 3) que los dos pueblos sean los extremos de sus diámetros. En realidad, el reactivo no pedía que se cumpliera la tercera condición; sin embargo, el hecho de que en la figura del pie del ítem se haya incluido una línea para ilustrar la distancia entre ambas localidades, al parecer fue lo que motivó esa confusión generalizada, de la cual hubo muchas evidencias.

La siguiente verbalización de uno de los alumnos de NSE alto ilustra el problema cuando explica porqué seleccionó la opción de ninguna circunferencia y dice:

porque si tengo 4 kilómetros de radio el diámetro sería de 8 entonces [señalando la línea de la Figura 27 que representa la distancia entre los dos pueblos, como se puede ver en la circunferencia superior de la siguiente figura que trazó] queda entre el círculo y queda pequeño entonces no alcanza a tomar [mientras traza en el aire una circunferencia] (E08).

Luego cuando se le sondeó qué parte del reactivo fue lo que le hizo que pensara que la circunferencia debería tener el acomodo tal y como la dibujó en la circunferencia superior de la figura, el alumno responde:

que dice “al mismo tiempo”. Puede que el círculo esté cargado así [mientras dibuja la circunferencia inferior de la figura], pero dice “al mismo tiempo” entonces aquí no, por lo tanto es ninguna (E08).



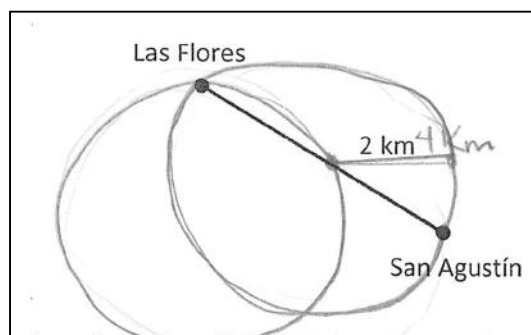
**Figura 27. Evidencia del informante E08 para responder el ítem SMH-11**

Por su parte, la alumna de NSE bajo también parece comprender que los puntos deben ser los extremos del mismo diámetro, pues cuando selecciona la opción de respuesta d) que es “ninguna” y se le pregunta por qué eligió esa opción, responde:

dice “un radio de 4”. Entonces si hago un círculo y el radio es del centro hacia cualquiera de los lados de la superficie, y dice que de cuatro, pero la distancia es de 2, entonces sería más cortito (E02).

Parte del proceso que siguió la alumna de NSE medio para resolver el reactivo consistió en trazar un círculo cuyos extremos del diámetro son la ubicación de los dos pueblos, como se muestra en la Figura 28. La alumna señala que el círculo (derecho en la figura) mide 4 km de radio; vale decir que esta circunferencia no es plausible porque si tuviera un radio de 4 km, las dos localidades no podrían ser los extremos de su diámetro, tal y como ella la dibujó. Luego la entrevistada traza otro

círculo (el izquierdo de la figura) en el que la circunferencia pasa por Las Flores y el punto medio entre los dos pueblos, y la alumna verbaliza: “pero éste ya no pasaría por San Agustín, entonces nada más sería una...[circunferencia]” (E04).



**Figura 28. Evidencia del informante E04 para responder el ítem SMH-11**

Nótese cómo en la Figura 28, las dos localidades se consideraron como extremos del diámetro de una de las circunferencias, y esa fue la confusión principal del reactivo, que se sospecha fue ocasionado por la inclusión de la línea que intenta ilustrar una distancia de 2 km. Así, cuando la entrevistadora le pregunta al docente que atiende alumnos de NSE alto “¿no puede haber una circunferencia de cuatro kilómetros que pase por esos dos puntos? [señalando los puntos que ubican a las dos localidades]”, el profesor responde:

no, porque la distancia son dos. Una circunferencia de cuatro viene siendo hasta acá [al tiempo que proyecta un punto siguiendo la misma línea que ilustra la distancia de los dos 2 km.] (E09).

Además, otra muestra del problema de ilegibilidad del ítem es que al menos cinco entrevistados tuvieron dificultades para comprender la pregunta del pie del ítem. Por ejemplo, uno de los alumnos de NSE bajo señala que el pie del reactivo le

parece confuso, y cuando se le pide que explique qué es lo que le causa confusión, verbaliza la pregunta del pie del ítem: “¿Cuántas circunferencias de 4 kilómetros de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?” (E01).

Otro de los alumnos entrevistados (NSE alto) después de leer en una primera ocasión el pie del reactivo, dice:

Ok. Aquí lo voy a leer otra vez porque no entendí muy bien a la pregunta. Esto sí está muy claro [refiriéndose a la figura]. No sé muy bien si al decir circunferencia se refiere a que sean círculos, cuántos círculos pasan así [mientras con su lápiz forma círculos imaginarios sobre la figura del pie del ítem, de tal forma que San Agustín y Las Flores sean extremos de su diámetro], o si del círculo, que sea como un hilo que se estire, cuántas veces son... (E07).

Este mismo alumno comprende que la línea del pie del ítem que ilustra la distancia de 2 km entre los dos pueblos debe ser parte del diámetro de las posibles circunferencias, y que por tanto ninguna circunferencia pasa por ambas localidades, cuando verbaliza:

Ok. Entonces tengo que pensarlo de esta manera. De cualquier manera no cabría, porque ya cuatro kilómetros de radio ya sería el doble de esto [mientras simula una línea de 4 km sobre la línea de la figura]... no cabe ningún círculo ahí (E07).

Uno de los docentes (de NSE alto) lee en tres ocasiones el pie del reactivo, en especial la pregunta “¿Cuántas circunferencias de 4 km de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?” y verbaliza:



Eso sí no entiendo. No le entiendo. Aquí son dos kilómetros [señalando la línea de la figura del pie del ítem que ilustra la distancia entre los dos pueblos]. No, no es comprensible la pregunta (E09).

Cuando la entrevistadora le sondea porqué no se comprende, el docente responde:

porque este dato [señalando la línea de la figura del pie del ítem que ilustra la distancia entre los dos pueblos] no sé qué tiene que ver con los 4 kilómetros de radio de la circunferencia (E09).

En resumen, la evidencia apunta a un problema de incomprensibilidad o ilegibilidad, muy probablemente ocasionado por la línea que se utilizó para ilustrar la distancia entre los pueblos, que en lugar de tener ese significado entre los sustentantes, fungió en el marco del ítem como una indicación de que los círculos deberían ser concéntricos, cuando en realidad el reactivo no lo planteaba de esa manera.

#### **5.4.2. Incomprensibilidad o problemas de legibilidad del ítem SME-09**

El ítem SME-09 (ver Figura 29) pertenece al eje de la asignatura de Matemáticas llamado *manejo de la información*; en el programa de Matemáticas corresponde al tema *representación de la información* y al subtema *diagramas y tablas*. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas de conteo que se puedan resolver por medio de una multiplicación.

¿De cuántas formas se pueden obtener números de cuatro cifras con los dígitos 1, 2, 3, 4, 5, 6, 7, 8, 9 y 0? Considera que los dígitos se pueden repetir.

$10 \times 9 \times 8 \times 7$

$10 \times 8 \times 6 \times 4$

$10 \times 9 \times \dots \times 1$

$9 \times 10 \times 10 \times 10$  \*

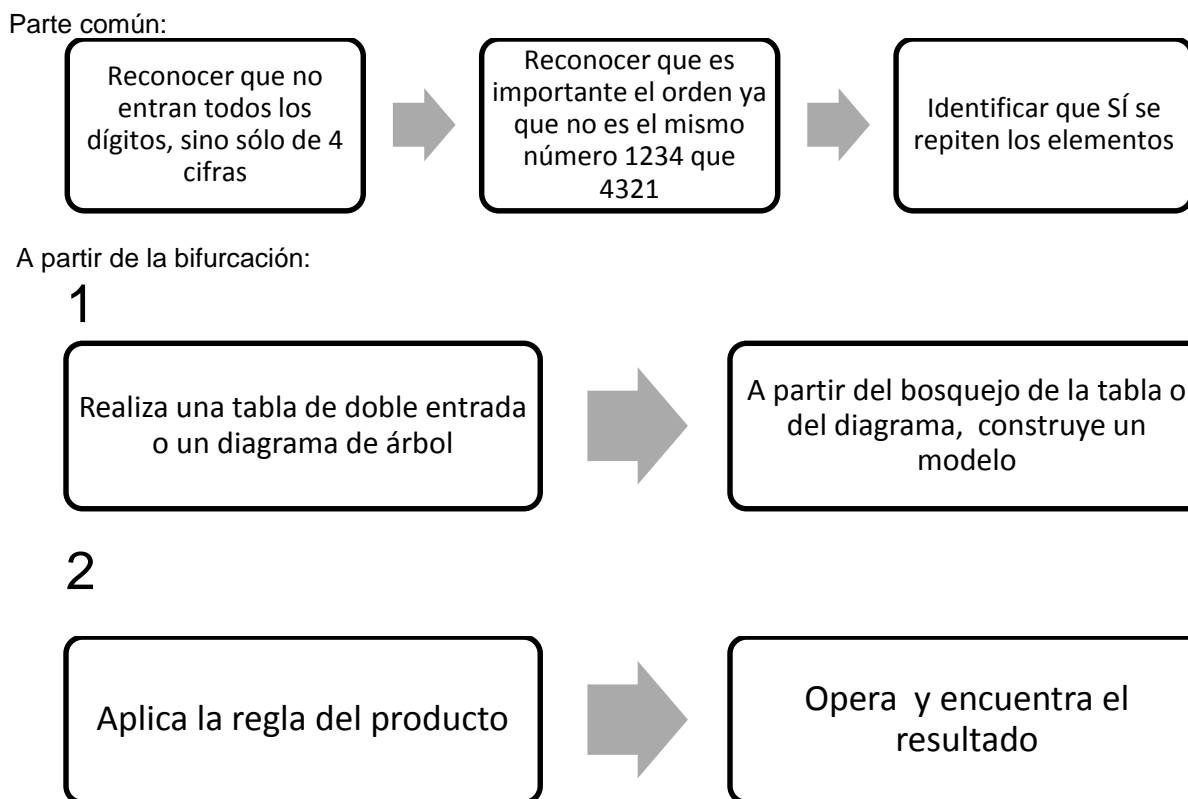
\* Opción correcta

**Figura 29. Reactivo SME-09**

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Problemas de conteo y sus diferentes representaciones.
- Uso de diagramas de árbol y tablas de doble entrada.
- Conocer y comprender la regla del producto.
- Analizar las diferentes posibilidades planteadas.
- Identificar el producto que resuelve el problema (INEE, 2010f, p. 7).

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los dos senderos cognitivos que se ilustran en el siguiente esquema, los cuales tienen una parte común (ver Figura 30):



**Figura 30. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-09**

El análisis de las videograbaciones de los protocolos y reportes verbales aportan elementos para afirmar sobre un posible problema de incomprendibilidad o legibilidad del ítem. Cuatro de los entrevistados (los tres docentes y un alumno de NSE bajo) aportaron elementos en ese sentido. El docente de NSE bajo, quien respondió correctamente el ítem, al ir verbalizando el proceso cognitivo que empleaba, dice:

Veo un problema en ésta [pregunta]. No les dice que el primer dígito no puede ser cero. Dice ‘considera números de cuatro dígitos’ y entonces aquí, al elegir el primer número, el primer dígito, son diez opciones, al escribir el segundo tienen otras 10 opciones, el tercero otras 10 opciones y

el cuarto otras 10 opciones... y ninguna de las respuestas probables representa esto (E03).

Luego la entrevistadora lo sondea con “¿y entonces cómo la razonó usted?”, a lo que el docente responde “el primer dígito no puede ser cero, entonces nada más tengo nueve opciones para el primero” (E03). La entrevistadora plantea una nueva pregunta “¿y considera que esa sería una dificultad?”, a lo que el profesor responde “esa sería una dificultad para el alumno, creo que los niños no lo relacionan de esa manera, entonces tendría que decir que el primer dígito no puede ser cero” (E03). Además, el docente argumenta que se podría escribir “0123” y sería un número de 4 cifras, que se sabe que el cero no tomaría valor, pero que sigue siendo un número de cuatro dígitos que es lo que plantea el problema.

Los otros dos docentes se sorprendieron cuando bajo su razonamiento la respuesta correcta debería ser  $10 \times 10 \times 10 \times 10$  y no la encontraron dentro de las opciones. Es decir, si bien los docentes sabían que el problema implicaba una permutación donde los dígitos se puede repetir, el reactivo tiene un elemento de complejidad innecesaria, más asociado a la Lógica y que no forma parte de lo que el reactivo pretende medir, en la que los sustentantes deben considerar que en el conjunto de los números de cuatro cifras, aquellos que inicien con 0 no suele considerarse como parte de ese conjunto.

De acuerdo con el DUE y con los estándares de medición para pruebas educativas (AERA-APA-NCME, 2014; Thompson *et al.*, 2002), un reactivo sólo debería evaluar aquello que contribuya a medir el constructo que se pretende

evaluar. El resto de componentes que se incluyan al reactivo, si involucran habilidades o conocimientos a los que son objetivo de evaluación, deberían de ser desechados y, por el contrario, es permisible y deseable incorporar acomodaciones para subsanar los obstáculos a que son expuestas ciertas subpoblaciones de sustentantes (que en este caso son los alumnos según su NSE) sin que se altere las habilidades que ellos tienen que emplear para resolver las tareas evaluativas, de tal manera que los ponga en ventaja con respecto al resto de evaluados.

## CONCLUSIONES

Una primera etapa de esta investigación doctoral consistió en el análisis de cada uno de los cien ítems del Excale de matemáticas para secundaria mediante la técnica RASCH-DIF y la utilización de dos programas estadísticos (Winsteps v. 3.91 y R Statistical v. 3.3.1) que permitió corroborar la concurrencia del diagnóstico. Los resultados de este análisis constataron que los ítems del Excale están libres de sesgo por sexo; en cambio, se identificaron 18 reactivos con DIF severo según el nivel socioeconómico (NSE), de los cuales ocho resultaron más difíciles para los alumnos de NSE alto: SMB\_01, SMF\_02, SMB\_09, SMG\_07, SMB\_11, SMC\_08, SME\_11 y SME\_12; otros ocho ítems fueron más difíciles para estudiantes de NSE bajo: SMB\_03, SME\_02, SMD\_04, SMD\_05, SMH\_11, SMC\_10, SMD\_11 y SME\_09; asimismo, dos reactivos presentaron DIF No uniforme (NUDIF): SMA\_05, SMG\_05.

De la subescala Sentido Numérico y Pensamiento Algebraico (SNPA) se identificaron cuatro ítems con DIF severo (SMB\_01, SMB\_03, SME\_02 y SMF\_02) y dos con NUDIF severo (SMA\_05 y SMG\_05). En la subescala Forma, Espacio y Medida (FEM) se diagnosticaron cinco ítems con DIF severo (SMB\_09, SMD\_04, SMD\_05, SMG\_07 y SMH\_11) y ninguno con NUDIF. De la subescala Manejo de la Información (MI) se encontraron siete ítems con DIF severo (SMB\_11, SMC\_08, SMC\_10, SMD\_11, SME\_09, SME\_11 y SME\_12) y ninguno con NUDIF.

Que los ítems no presenten DIF por sexo es un hallazgo relevante, pues indica que los mecanismos que el Instituto Nacional para la Evaluación de la Educación (INEE) ha utilizado para cuidar que sus reactivos estén libres de sesgo a través de

la conformación de comités de validación, son efectivos. Sin embargo, la presencia de DIF por NSE también deja que ver que las estrategias que el INEE ha utilizado para minimizar el sesgo en estos reactivos no ha sido suficiente y, por el contrario, requiere la mejora progresiva y constante de procesos internos para el cuidado en el diseño de los ítems desde un enfoque de equidad en las evaluaciones.

Una vez identificados los ítems con DIF severo, se exploró a través de reportes y protocolos verbales, la presencia de sesgo potencial en esos 18 ítems del Excale 09 de Matemáticas. Para llevar a cabo dicho análisis se utilizaron cuatro de las siete categorías del Diseño Universal de las Evaluaciones (DUE) para analizar la información recolectada, a saber: 1) definición imprecisa del constructo, 2) ítem no accesible o sesgado, 3) procedimientos e instrucciones complejas, poco claras e intuitivas y 4) incomprendibilidad o problemas de legibilidad del ítem.

Los resultados pueden sintetizarse de la siguiente manera: se identificaron 13 ítems con sesgo en al menos una de las cuatro categorías del DUE a partir de las cuales se les analizó. En la subescala MI es donde se encontraron más ítems (seis) con sesgo potencial, en cambio en la subescala SNPA se diagnosticaron tres reactivos con riesgo de sesgo y en la subescala FEM se identificaron cuatro ítems con una alta posibilidad de sesgo. El tipo de sesgo que más se presentó fue la incomprendibilidad o problemas de legibilidad en ocho ítems, en cambio, ninguno de los reactivos tuvo evidencias de que los procedimientos e instrucciones que se les ofrecieron a los sustentantes para saber cómo registrar sus respuestas fueran complejas, poco claras e intuitivas (ver Tabla 21).

**Tabla 21. Cantidad de informantes que aportaron evidencias en los ítems diagnosticados con sesgo potencial, desglosado por subescala y tipo de sesgo**

Subescala	Folio Reactivo	Tipos de sesgo			
		Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
SNPA	SMB-01	0	4	0	3
	SMB-03	3	0	0	0
	SME-02	3	0	0	0
FEM	SMB-09	0	0	0	2
	SMD-04	0	7	0	0
	SMG-07	0	1	0	0
	SMH-11	0	0	0	7
MI	SMC-08	0	2	0	0
	SMC-10	8	3	0	4
	SMD-11	0	3	0	1
	SME-09	0	0	0	4
	SME-11	0	0	0	2
	SME-12	2	0	0	0
Núm. de reactivos con presencia de varianza irrelevante al constructo		4	6	0	7

**Notas:** SNPA = Sentido Numérico y Pensamiento Algebraico; FEM = Forma, Espacio y Medida; MI = Manejo de la Información

**Fuente:** elaboración propia.

En un ejercicio para integrar la información proveniente de los análisis DIF como de los protocolos y reportes verbales, en el que se ofrecen posibles explicaciones del sentido del DIF y las evidencias de sesgo, sirva la Tabla 22 para disponer de un panorama de la situación. Nótese que por cada reactivo se presenta el tipo de DIF con el que fue diagnosticado y la cantidad de informantes que aportaron evidencias de sesgo, desglosándolos por su NSE.



**Tabla 22. Cantidad de evidencias de sesgo, por reactivo, tipo de DIF, tipo de sesgo y nivel socioeconómico**

Subescala	Folio Reactivo	Tipo DIF	Definición imprecisa del constructo			Ítem no accesible o sesgado			Incomprensibilidad o problemas de legibilidad del ítem			Evidencias de sesgo
			B	M	A	B	M	A	B	M	A	
SNPA	SMB-01	++				2	2		1	1	1	7
	SMB-03	--	1	1	1							3
	SME-02	--	2		1							3
	SMF_02	++										0
	SMA_05	NU										0
	SMG_05	NU										0
FEM	SMD-04	--				3	3	1				7
	SMH-11	--							2	2	3	7
	SMB-09	++								1	1	2
	SMG-07	++				1						1
	SMD_05	--										0
MI	SMC-10	--	3	3	2	3				3	1	15
	SMD-11	--				2	1			1		4
	SME-09	--							2	1	1	4
	SME-12	++	1		1							2
	SMC-08	++					1	1				2
	SME-11	++							1	1		2
	SMB_11	++										0

**Notas:**

SNPA = Sentido Numérico y Pensamiento Algebraico; FEM = Forma, Espacio y Medida; MI = Manejo de la Información

++ DIF severo a favor de NSE Alto; -- DIF severo a favor de NSE Bajo; NU = No Uniforme

B = Nivel Socioeconómico Bajo; M = Nivel Socioeconómico Medio; A = Nivel Socioeconómico Alto

**Fuente:** elaboración propia

En términos particulares, el ítem SMB-01 (ver Figura 21), que presentó DIF severo a favor de NSE Alto, la evidencia disponible lleva a concluir que el reactivo incluye un término (descendió) que no forma parte del constructo que se pretende medir y que pone innecesariamente en desventaja a subgrupos de estudiantes de NSE Bajo, donde no es común la utilización de esa palabra y, como en el caso de uno de los entrevistados, los lleva a hacer interpretaciones opuestas del sentido

que ese vocablo tiene en realidad, lo que luego conlleva a que respondan de forma incorrecta el reactivo.

Y una situación similar ocurrió con el ítem SMD-04 (ver Figura 22), cuyo DIF favorece a estudiantes de NSE Bajo. Para resolverlo de forma correcta, resultaba indispensable conocer la equivalencia de una hectárea en metros cuadrados o kilómetros cuadrados. Considerando que existe una fuerte asociación entre NSE y ruralidad, es muy probable que para los sustentantes que provienen de medios rurales (y NSE Bajo), donde es más común la utilización de hectáreas como unidad para medir parcelas, terrenos o ranchos, disponían del conocimiento de la equivalencia; en cambio, ante la escasa o nula utilización de la hectárea como unidad medida en contextos urbanos (en los que suele haber una mayor proporción de estudiantes de NSE medio y alto), les dificultó innecesariamente la resolución del ítem.

La evidencia empírica recabada sobre estos dos ítems, hace patente cómo la utilización de un lenguaje que es extraño para ciertas subpoblaciones, puede en efecto, obstaculizar la resolución adecuada de las tareas evaluativas, a pesar de que dicho lenguaje no forma parte del objeto de evaluación, lo que constituye una fuente de sesgo.

En términos generales, la información de las tablas 21 y 22 llevan a preguntarse porqué para ciertos ítems se tienen evidencias de sesgo abundantes provenientes de la mayoría de los informantes, como el caso de los ítems SMC-10, SMD-04, SMH-11 y SMB-01, o bien, porqué existen otros reactivos de los cuales no se consiguieron evidencias de sesgo, tales como el SMG\_05, SMA\_05, SMD\_05,

SMB\_11 y SMF\_02. La respuesta inmediata conlleva a pensar que aquellos reactivos de los cuales se obtuvo evidencia de sesgo en realidad tengan algún tipo de inaccesibilidad o varianza irrelevante al constructo y, al mismo tiempo, que aquellos ítems en los cuales no se encontraron evidencias de sesgo en realidad están libres de él.

Otras posibles respuestas, circunscritas a los reactivos con DIF en los cuales no se encontró evidencia de sesgo mediante los protocolos verbales, son las siguientes: 1) Las tres categorías denominadas definición imprecisa del constructo, inaccesibilidad e incomprendibilidad están muy relacionadas con el contenido matemático que se evalúa; un eventual falso negativo en el diagnóstico podría ser rechazado a partir de la opinión de especialistas en enseñanza de las matemáticas, que ayuden a comprender los defectos de esos ítems identificados con DIF severo. 2) La sensibilidad a la instrucción<sup>9</sup> como una posible fuente que ayude a explicar el funcionamiento diferencial, sobre todo en reactivos de la subescala MI, que aunque no es lo recomendable por el currículo, se suelen tratar al final del ciclo; ciertos profesores que atienden alumnos de NSE bajo, podrían descartarlos por priorizar temáticas previas que no han consolidado sus alumnos.

Los resultados de esta investigación son de especial utilidad para los diseñadores de Excale (ahora PLANEA) porque permiten identificar reactivos en específico que conviene revisar, al mismo tiempo, ofrece una metodología que puede emplearse para la revisión del sesgo en reactivos de pruebas de logro o desempeño. La

---

<sup>9</sup> Dentro del campo del desarrollo de pruebas, se le denomina sensibilidad a la instrucción al fenómeno que explica por qué algunos ítems no son respondidos correctamente debido a que los contenidos que evalúan no fueron enseñados o tratados durante las clases.

combinación de técnicas de DIF con el uso concatenado de protocolos y reportes verbales permitió identificar, con un nivel inusual en los estudios del análisis de sesgo en México, fuentes de varianza irrelevante al constructo que pueden ser utilizadas para analizar otras pruebas que se apliquen de forma masiva para minimizar la inequidad en las evaluaciones.

Desde una visión retrospectiva y sin demeritar los logros que ya se han señalado, se reconoce que este trabajo tiene también algunas limitaciones. Se identifican las siguientes:

- En el diseño de los análisis DIF no se incluyó como variable de segmentación el tipo de lengua materna de los estudiantes, a pesar de que los resultados de logro muestran de forma consistente que en la modalidad indígena es donde se suelen tener, junto con las escuelas comunitarias, los niveles de logro más bajos.
- En el diseño de los análisis DIF tampoco se incluyó la modalidad educativa como una variable de comparación, porque se asumió que existe una fuerte asociación entre el NSE y esta variable.
- Para seleccionar a los estudiantes que participaron en los protocolos y reportes verbales se asumió que su NSE corresponde con el tipo de escuela secundaria en la que están adscritos. Así, a los alumnos de telesecundaria se les clasificó de NSE bajo, los que estaban inscritos a una secundaria general pública se les identificó de NSE medio y a los que estudiaban en una secundaria privada se les clasificó de NSE alto. Aunque estudios previos en el

contexto mexicano refieren la alta correlación entre la modalidad en la que se inscriben y el NSE de los alumnos, para este trabajo se trata de un supuesto.

- Aunque Excale ya no se aplica en el país, muchos de sus reactivos (incluidos los 18 que presentaron DIF severo) aún siguen vigentes en el banco de reactivos que se utiliza para diseñar las pruebas de PLANEA. Ante la imposibilidad de utilizar los cuadernillos originales en los protocolos verbales, y generar un cuadernillo *ad hoc* sólo con los reactivos que presentaron DIF, no se pudo incorporar como eje de análisis el elemento del DUE denominado máxima visibilidad.
- En los análisis de los protocolos y reportes verbales no se utilizaron los elementos del DUE llamados evaluación inclusiva de la población y sensibilidad a las adaptaciones. No se incluyeron porque requieren de metodologías que incluye conocer desde las etapas iniciales del diseño de las pruebas en las que, entre otras cuestiones, se define el tipo de población objeto de la evaluación, los estratos muestrales que se utilizan en las aplicaciones piloto y las adaptaciones que se van realizando a los reactivos para que sean más pertinentes a ciertas subpoblaciones, y que a la vez no incorporen fuentes de varianza irrelevante. El conocimiento y documentación de esos procesos rebasaron los alcances del estudio.

Desde una óptica prospectiva, conviene que los estudios para cuidar la equidad en las evaluaciones se sigan impulsando y fortaleciendo, lo que contribuirá a que las interpretaciones y usos que se hagan con los resultados de las pruebas mexicanas que se aplican a los alumnos de educación obligatoria sean

técnicamente más precisos y moralmente más justos. Con ese propósito, se ofrecen algunas recomendaciones:

1. Resulta necesario que en el desarrollo de los ítems, una vez piloteados, se sometan de manera sistemática a análisis DIF y, posteriormente, los ítems sospechosos de sesgo pase por algún mecanismo para descartar sesgo. Si bien pueden analizarse mediante protocolos verbales, esté método puede resultar costoso en tiempo y análisis de los resultados; por ello, los reactivos al menos tendrían que revisarse por grupos en que intervengan especialistas en enseñanza de los contenidos de que se trate, junto con maestros con experiencia con alumnos de sectores vulnerables, para detectar fallas como las que parece haber en algunos de los ítems analizados en este trabajo.
2. Disponer de evaluaciones equitativas implica la incorporación de una serie de mecanismos con un propósito explícito en ese sentido. La recomendación es que dichos mecanismos sean incluidos desde las fases iniciales en que diseñan las evaluaciones. Resulta muy conveniente la incorporación de ciertos perfiles (como antropólogos, lingüistas y sociólogos) en los equipos que coordinan la definición operacional de los constructos a evaluar, el diseño de las especificaciones, los comités de validación de ítems, el diseño muestral y la interpretación de los resultados, todo ello con el propósito de valorar la pertinencia y adecuación de esas evaluaciones a las diferentes subpoblaciones que son su población objetivo, y hacer los ajustes necesarios para que las subpoblaciones de interés queden debidamente consideradas en la manera que se utilizarán los resultados.

3. Como parte del procedimiento habitual del diseño de las evaluaciones se sugiere incorporar la realización de ciertos análisis psicométricos que busquen identificar de forma oportuna varianza irrelevante al constructo a través de las técnicas que resulten más apropiadas; como se ha visto en esta tesis, existe una gran variedad de herramientas para la detección de DIF, cada una con su ventajas y desventajas, por lo que conviene utilizar más de una para robustecer y complementar el diagnóstico.
4. En la actualidad siguen desarrollándose nuevas técnicas para la detección de DIF tales como FACET-DIF, DIF por medio del Modelo lineal de test logístico (LLTM en inglés), DIF a través del modelo DINA, etc. Sería muy conveniente ir incorporándolas, según las posibilidades del organismo diseñador de las evaluaciones. No obstante, además de los procedimientos psicométricos para el cuidado en la equidad de las evaluaciones conviene incluir otro tipo de técnicas para identificar varianza irrelevante al constructo, tal como los protocolos verbales, los reportes verbales, entre otras. Cada una de ellas aporta distintos tipos de evidencias de validez, que en su conjunto, fortalecerían la validez en la interpretación y uso de los resultados.
5. Se sugiere que, en los análisis de sesgo, de manera gradual se vayan incorporando como eje de análisis las subpoblaciones indígenas más numerosas del país, lo que favorecerá la realización de evaluaciones que guarden el principio de equidad, y que se tenga más probabilidades de que los sustentantes de estas modalidades tengan la oportunidad de mostrar sus talentos a cabalidad cuando se enfrentan a las tareas evaluativas.

## REFERENCIAS

- AERA, APA y NCME. (2014). *Standards for educational and psychological testing*. Washington: AERA-APA-NCME.
- AERA-NCME-APA. (1999). *Standards for Educational and Psychological Testing*. Washington: AERA-APA-NCME.
- Backhoff, E., Bouzas, A., González-Montesinos, M., Andrade, E. y Hernández, E. (2008). *Factores asociados al aprendizaje de estudiantes de 3º de primaria en México*. Ciudad de México: INEE.
- Backhoff, E., Peón, M. y Sánchez, A. (2009). *Manual Técnico para el Diseño de Exámenes de la Calidad y el Logro Educativos (2a ed.)*. Ciudad de México: INEE.
- Backhoff, E., Peón, M., Sánchez, A. y Andrade, E. (2006). *Excale. Manual Técnico para la Validación de reactivos*. Ciudad de México: INEE.
- Backhoff, E., Solano-Flores, G., Contreras, L. Á., Vázquez, M. y Sánchez, A. (2015). *Son adecuadas las traducciones para evaluar los aprendizajes de los estudiantes indígenas. Un estudio con preescolares mayas*. Ciudad de México: INEE.
- Balluerka, N., Gorostiaga, A., Gómez-Benito, J. y Hidalgo, M. D. (2010). Use of multilevel logistic regression to identify the causes of differential item functioning. *Psicothema*, 22(4), 1018–1025.
- Basterra, M. del R., Trumbull, E. y Solano-Flores, G. (2011). *Cultural Validity in Assessment*. New York: Routledge.



- Benítez, I. y Padilla, J. L. (2014). Analysis of Nonequivalent Assessments Across Different Linguistic Groups Using a Mixed Methods Approach Understanding the Causes of Differential Item Functioning by Cognitive Interviewing. *Journal of Mixed Methods Research*, 8(1), 52–68. Recuperado en: <https://doi.org/10.1177/1558689813488245>
- Blanco, E. E. (2007). *Eficacia escolar en México: factores escolares asociados a los aprendizajes en la educación primaria*. Ciudad de México: FLACSO. Recuperado en: <http://flacsoandes.edu.ec/dspace/handle/10469/1247>
- Boone, W. J., Staver, J. R. y Yale, M. S. (2014). Differential Item Functioning. En *Rasch Analysis in the Human Sciences*, pp. 273–297. Netherlands: Springer. Recuperado en: [http://link.springer.com.dibpxy.uaa.mx/chapter/10.1007/978-94-007-6857-4\\_13](http://link.springer.com.dibpxy.uaa.mx/chapter/10.1007/978-94-007-6857-4_13)
- Borsboom, D., Mellenbergh, G. J. y Van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061–1071.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition). New York: The Guilford Press.
- Byrne, B. M., Shavelson, R. J. y Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological bulletin*, 105(3), 456.
- Camilli, G. (1992). A Conceptual Analysis of Differential Item Functioning in Terms of a Multidimensional Item Response Model. *Applied Psychological Measurement*, 16(2), 129–147. Recuperado en: <https://doi.org/10.1177/014662169201600203>

- Camilli, G. (2006). Test Fairness. En R. L. Brennan (Ed.), *Educational Measurement*, 4a ed.. Washington: ACE-NCME, pp. 221–256.
- Camilli, G. y Shepard, L. A. coautor. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage Publications.
- Castillo-Díaz, M. y Padilla, J.-L. (2013). How Cognitive Interviewing can Provide Validity Evidence of the Response Processes to Scale Items. *Social Indicators Research*, 114(3), 963–975. Recuperado en: <https://doi.org/10.1007/s11205-012-0184-8>
- Cea D' Ancona, M. Á. (2001). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Madrid: Síntesis Sociología.
- Center for Universal Design. (1997). *Los Principios Del Diseño Universal*. Center for Universal Design, North Carolina State University. Recuperado en: [https://www.ncsu.edu/ncsu/design/cud/about\\_ud/docs/Spanish.pdf](https://www.ncsu.edu/ncsu/design/cud/about_ud/docs/Spanish.pdf)
- Chun, S., Stark, S., Kim, E. S. y Chernyshenko, O. S. (2016). MIMIC Methods for Detecting DIF Among Multiple Groups: Exploring a New Sequential-Free Baseline Procedure. *Applied Psychological Measurement*, 40(7), 486–499. Recuperado en: <https://doi.org/10.1177/0146621616659738>
- Clauser, B. E. y Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice*, 17(1), 31–44. Recuperado en: <https://doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Clauser, B., Mazor, K. M. y Hambleton, R. K. (1994). The Effects of Score Group Width on the Mantel-Haenszel Procedure. *Journal of Educational*

- Measurement*, 31(1), 67–78. Recuperado en: <https://doi.org/10.1111/j.1745-3984.1994.tb00435.x>
- Cole, N. S. y Moss, P. A. (1989). Bias in Test Use. En R. L. Linn, *Educational Measurement*, 3a ed. New York: ACE-NCME, pp. 201–219.
- Cole, N. S. y Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369–382. Recuperado en: <https://doi.org/10.1111/j.1745-3984.2001.tb01132.x>
- Cronbach, L. J. (1971). Test Validation. En R. Thorndike (Ed.), *Educational Measurement*, 2a ed. Washington: American Council on Education, pp. 443–507.
- Cureton, E. E. (1951). Validity. En E. Linn, *Educational Measurement*. Washington: American Council on Education, pp. 621–294.
- Engelhard, G. (1992). The Measurement of Writing Ability With a Many-Faceted Rasch Model. *Applied Measurement in Education*, 5(3), 171–191. Recuperado en: [https://doi.org/10.1207/s15324818ame0503\\_1](https://doi.org/10.1207/s15324818ame0503_1)
- Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F. y Lacroix, S. (2010). Application of Think Aloud Protocols for Examining and Confirming Sources of Differential Item Functioning Identified by Expert Reviews. *Educational Measurement: Issues and Practice*, 29(2), 24–35. Recuperado en: <https://doi.org/10.1111/j.1745-3992.2010.00173.x>
- Ericsson, K. A. y Simon, H. A. (1993). *Protocol analysis. Verbal Reports as Data* 2th ed. Cambridge: MIT Press. Recuperado en: <http://ammonwiemers.com/ldetPortfolio/articles/Assessment/Protocol%20Analysis%20--%20Verbal%20Reports%20as%20Data.pdf>

- ETS. (2002). *ETS Standards for Quality and Fairness*. Princeton: Educational Testing Service.
- ETS. (2009). *ETS Guidelines for Fairness Review of Assessments*. Princeton: Educational Testing Service. Recuperado en: [https://www.ets.org/Media/About\\_ETS/pdf/overview.pdf](https://www.ets.org/Media/About_ETS/pdf/overview.pdf)
- ETS. (2014). *ETS Standards for Quality and Fairness*. Princeton: Educational Testing Service.
- Fidalgo, A. M. (1997). Funcionamiento Diferencial de los Ítems. En J. Muñiz, *Psicometría*. Madrid: Editorial Universitas, pp. 372–455.
- Fleck, J. y Weisberg, R. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory y Cognition*, 32(6), 990–1006. Recuperado en: <https://doi.org/10.3758/BF03196876>
- García, J. L. P., Gómez-Benito, J. y Hidalgo, M. D. (2005). Regresión logística: alternativas de análisis en la detección del funcionamiento diferencial del ítem. *Psicothema*, 17(3), 509–515.
- Gipps, C. y Stobart, G. (2009). Fairness in Assessment. En C. Wyatt-Smith y J. J. Cumming (Eds.), *Educational Assessment in the 21st Century*. Dordrecht: Springer Netherlands, pp. 105–118. Recuperado en: <http://link.springer.com/10.1007/978-1-4020-9964-9>
- Gipps, C. y Stobart, G. (2010). Fairness. En *International Encyclopedia of Education*, 3a ed., Vol. 4. Oxford: Elsevier, pp. 56–60.
- González-Montesinos, M. (2008). El Análisis de Reactivos con el Modelo Rasch. Recuperado en: <http://www.winsteps.com/a/recursos-offline.pdf>

- González-Montesinos, M. y Jornet, J. M. (2012). *Procedimientos para la Detección del Funcionamiento Diferencial de Reactivos (DIF)*. Documento preparado para el INEE.
- Hernández, A. y González-Romá, V. (2003). Evaluating the Multiple-Group Mean and Covariance Structure Analysis model for the detection of Differential Item Functioning in polytomous ordered items. *Psicothema*, 15(2), 322–327.
- Hidalgo, M. D. y Gómez-Benito, J. (2010). Differential Item Functioning. En P. P. B. McGaw (Ed.), *International Encyclopedia of Education*. 3a ed. Vol. 4. Oxford: Elsevier, pp. 36–44.
- Hidalgo, M. D., Gómez-Benito, J. y Zumbo, B. D. (2014). Binary Logistic Regression Analysis for Detecting Differential Item Functioning Effectiveness of  $R^2$  and Delta Log Odds Ratio Effect Size Measures. *Educational and Psychological Measurement*, 74(6), 927–949. Recuperado en: <https://doi.org/10.1177/0013164414523618>
- Holland, P. W. y Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. En H. Wainer y H. I. Braun, *Test validity*. Hilldale: L. Erlbaum, pp. 129–145.
- INEE. (2010a). *Especificaciones de los reactivos del Excale-09*. Documento interno. INEE.
- INEE. (2010b). *Excale 09. Especificación de reactivos M0910ABA007*. Documento interno.
- INEE. (2010c). *Excale 09. Especificación de reactivos M0910ABC017*. Documento interno.

- INEE. (2010d). *Excale 09. Especificación de reactivos M0910BAA041*. Documento interno.
- INEE. (2010e). *Excale 09. Especificación de reactivos M0910BBA058*. Documento interno.
- INEE. (2010f). *Excale 09. Especificación de reactivos M0910CBA083*. Documento interno.
- INEE. (2010g). *Excale 09. Especificación de reactivos M0910CBB091*. Documento interno.
- Johnstone, C. (2003). *Improving Validity of Large-Scale Tests: Universal Design and Student Performance*. NCEO Technical Report. Recuperado en: <http://conservancy.umn.edu/handle/11299/174026>
- Juárez, E., Ramírez, R. y Rodríguez, G. (2006). *Manual Técnico para el Muestreo Poblacional. Excale*. Ciudad de México: INEE.
- Kane, M. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342.
- Kane, M. (2006). Validation. En R. L. Brennan (Ed.), *Educational Measurement*, 4a ed. Washington: ACE-NCME, pp. 17–64.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kiefer, T., Robitzsch, A. y Wu, M. (2016). *TAM: Test analysis modules*. Recuperado en: <http://www.edmeasurementsurveys.com/TAM/Tutorials/>
- Lee, S., Bulut, O. y Suh, Y. (2016). Multidimensional Extension of Multiple Indicators Multiple Causes Models to Detect DIF. *Educational and*

*Psychological Measurement*. Recuperado en:

<https://doi.org/10.1177/0013164416651116>

Lei, P.-W. y Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, 26(3), 33–43.

Leighton, J. P. (2009). Two Types of Think Aloud Interviews for Educational Measurement. Presentado en *National Council on Measurement in Education*, San Diego, CA. Recuperado en: [http://www2.education.ualberta.ca/educ/psych/crame/files/NCME2009\\_Two\\_Types\\_of\\_Think-Aloud\\_Interviews\\_for\\_Educational\\_Measurement.pdf](http://www2.education.ualberta.ca/educ/psych/crame/files/NCME2009_Two_Types_of_Think-Aloud_Interviews_for_Educational_Measurement.pdf)

Lissitz, R. W. (Ed.). (2009). *The concept of validity. Revisions, New Directions and Applications*. Charlotte: Information Age Publ.

Lissitz, R. W. y Samuelson, K. (2007). A Suggested Change in Terminology and Emphasis Regarding Validity and Education. *Educational Researcher*, 36(8), 437–448. Recuperado en: <https://doi.org/10.3102/0013189X07311286>

Martínez-Rizo, F. (2016). *La evaluación de docentes de educación básica. Una revisión de la experiencia internacional*. Ciudad de México: INEE. Recuperado en: <http://publicaciones.inee.edu.mx/buscadorPub/P1/C/233/P1C233.pdf>

Martínez-Rizo, F. y Blanco, E. (2010). La Evaluación Educativa en México. Experiencias, avances y desafíos. En A. Arnaut y S. Giorguli (Eds.), *Educación Mexicana. Situación actual y perspectiva*. Ciudad de México: Colegio de México. Recuperado en:

[http://www.fmrizo.net/fmrizo\\_pdfs/capitulos/C%20047%202010%20Evaluacion%20Educativa%20en%20Mexico\\_FMR-EB%20COLMEX.pdf](http://www.fmrizo.net/fmrizo_pdfs/capitulos/C%20047%202010%20Evaluacion%20Educativa%20en%20Mexico_FMR-EB%20COLMEX.pdf)

Martínez-Rizo, F., Contreras, L. Á., González, E., Jornet, J. M., Martínez, M. R., Martínez, J. F., Tristán, A. (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación*. Ciudad de México: INEE.

Martínez-Rizo, F. y Santos, A. (2009). Evaluación e investigación educativa. En A. Alicia (Ed.), *¿Qué dice la investigación educativa?*. Ciudad de México: COMIE, pp. 265–304. Recuperado en: [http://www.riieeme.mx/docs/docs%202013/Martinez%20Rizo\\_2009%20Evaluacion%20educativa.pdf](http://www.riieeme.mx/docs/docs%202013/Martinez%20Rizo_2009%20Evaluacion%20educativa.pdf)

McNamara, T. y Roever, C. (2006). Psychometric Approaches to Fairness: Bias and DIF. *Language Learning*, 56(S2), pp. 81–128. Recuperado en: <https://doi.org/10.1111/j.1467-9922.2006.00381.x>

Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational Measurement*, 3a ed New York: ACE-NCME, pp. 13–103.

Miller, K., Willson, S., Chepp, V. y Padilla, J. L. (Eds.). (2014). *Cognitive interviewing methodology*. Hoboken, New Jersey: John Wiley y Sons

Moss, P. A. (1992). Shifting conceptions of validity in educational measurements: Implications for performance assessments. *Review of Educational Research*, (62), 229–258.

Newton, P. (2013, febrero). *Does it matter what “validity” means?* Presentado en Seminar University of Oxford, Department of Education, Oxford.

Newton, P. y Shaw, S. (2014). *Validity in Educational and Psychological Assessment*. SAGE.



- OECD. (2009). *PISA 2006 Technical Report*. Recuperado en: <https://doi.org/http://dx.doi.org/10.1787/9789264048096-en>
- OECD. (2012). *PISA 2009 Technical Report*. Recuperado en: <https://doi.org/http://dx.doi.org/10.1787/9789264167872-en>
- Padilla, J.-L. y Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1). Recuperado en: <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=02149915&AN=93818805&h=pa97BtDHHHKXtgqrkCRvntq6HDEiutcr1769R3GQmAgkCzjpLfzwxGOFpN9SC2U0dhlpFfl%2BhQKc%2Fs99C1hVz7A%3D%3D&crl=c>
- Ployhart, R. E. y Oswald, F. L. (2004). Applications of Mean and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches. *Organizational Research Methods*, 7(1), 27–65. Recuperado en: <https://doi.org/10.1177/1094428103259554>
- Reise, S. P., Widaman, K. F. y Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological bulletin*, 114(3), 552.
- Reynolds, C. R. y Suzuki, L. (2013). Bias in Psychological Assessment. An Empirical Review and Recommendations. En I. B. Weiner (Ed.), *Handbook of Psychology*, 2nd ed. Vol. 10. John Wiley y Sons, pp. 82–113.
- Robles, H., Hernández, J. M., Zendejas, L. E. y Pérez, M. G. (2013). *Panorama Educativo de México 2012. Indicadores del Sistema Educativo Nacional. Educación Básica y Media Superior*. Ciudad de México: INEE.

- Ruiz, G., Pérez, M. G., Langford, P. y García-Medina, A. M. (2015). *Atención a la diversidad en evaluaciones educativas externas. Muestra de prácticas internacionales*. Ciudad de México: INEE.
- Sánchez, A. y Andrade, E. (2009). *El aprendizaje en tercero de secundaria en México. Informe sobre los resultados del Excale 09, aplicación 2008 Español, Matemáticas, Biología y Formación cívica y ética*. Ciudad de México: INEE.
- Sánchez, A., Andrade, E., Estrada, E., García, M., Guevara, G. y Rodríguez, G. (2012). *Estudio comparativo del aprendizaje en tercero de secundaria en México 2005-2008. Español y Matemáticas*. Ciudad de México: INEE.
- Sharma, S., Durvasula, S. y Ployhart, R. E. (2012). The Analysis of Mean Differences Using Mean and Covariance Structure Analysis Effect Size Estimation and Error Rates. *Organizational Research Methods*, 15(1), 75–102. Recuperado en: <https://doi.org/10.1177/1094428111403154>
- Shavelson, R. J. y Webb, N. M. (1991). *Generalizability Theory. A Primer*. Thousand Oaks: SAGE.
- Shavelson, R. J. y Webb, N. M. (2006). Generalizability theory. *Handbook of complementary methods in education research*, pp. 309–322.
- Shealy, R. y Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159–194. Recuperado en: <https://doi.org/10.1007/BF02294572>

- Shepard, L. A., Camilli, G. y Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 77–105.
- Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices. En *Cultural validity in assessment*. New York: Routledge, pp. 3–21.
- Solano-Flores, G. y Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573.
- Svetina, D. y Rutkowski, L. (2014). Detecting differential item functioning using generalized logistic regression in the context of large-scale assessments. *Large-Scale Assessments in Education*, 2(1), 1–17. Recuperado en: <https://doi.org/10.1186/s40536-014-0004-5>
- Thompson, S., Johnstone, C. y Thurlow, M. L. (2002). *Universal design applied to large scale assessments*. National Center on Educational Outcomes Synthesis Report. Recuperado en: [https://osepideasthatwork.org/Toolkit/pdf/Universal\\_Design\\_LSA.pdf](https://osepideasthatwork.org/Toolkit/pdf/Universal_Design_LSA.pdf)
- Tourangeau, R. (1984). Cognitive sciences and survey methods. En T. Jabine, M. Straf, J. Tanur y R. Tourangeau, *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington: National Academy Press, pp. 73–100.
- Treviño, E., Place, K., Gempp, R. y Donoso, F. (2013). *Análisis de los factores latentes y su vínculo con los resultados académicos de los niños*. Santiago, Chile: OREALC/UNESCO. Recuperado en: <http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Santiago/pdf/factores-asociados-al-aprendizaje-en-el-serce.pdf>

- Willis, G. B. (2005). *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks: Sage.
- Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford: Oxford University Press.
- Xie, Y., y Wilson, M. (2008). Investigating DIF and extensions using an LLTM approach and also an individual differences approach: an international testing context. *Psychology Science*, 50(3), 403.
- Zamudio, C., Díaz Argüero, C. y Lepe García, E. (2012). *El aprendizaje de los contenidos curriculares de español un análisis de los resultados de los Exámenes de la Calidad y el Logro Educativos (Excale 03, 06 y 09)*. Ciudad de México: INEE.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense. Recuperado en: <http://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf>
- Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement* (ETS. Research Report No. 8). ETS. Recuperado en: <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2012.tb02290.x/abstract>

## **ANEXOS**

**Anexo 1 Reactivos por cada forma de Cuadernillo del Excale 09 de Matemáticas. Aplicación 2012**

FOLIO REACTIVO	IDENTIFICADOR DE CUADERNILLO							
	PN2012SE C3ALF08	PN2012SE C3ALF09	PN2012SE C3ALF10	PN2012SE C3ALF11	PN2012SE C3ALF12	PN2012SE C3ALF13	PN2012SE C3ALF14	PN2012SE C3ALF15
SMA_01	1					1		1
SMA_02	1					1		1
SMA_03	1					1		1
SMA_04	1					1		1
SMA_05	1					1		1
SMA_06	1					1		1
SMA_07	1					1		1
SMA_08	1					1		1
SMA_09	1					1		1
SMA_10	1					1		1
SMA_11	1					1		1
SMA_12	1					1		1
SMB_01	1	1					1	
SMB_02	1	1					1	
SMB_03	1	1					1	
SMB_04	1	1					1	
SMB_05	1	1					1	
SMB_06	1	1					1	
SMB_07	1	1					1	
SMB_08	1	1					1	
SMB_09	1	1					1	
SMB_10	1	1					1	
SMB_11	1	1					1	
SMB_12	1	1					1	
SMB_13	1	1					1	
SMC_01		1	1					1
SMC_02		1	1					1
SMC_03		1	1					1
SMC_04		1	1					1
SMC_05		1	1					1
SMC_06		1	1					1
SMC_07		1	1					1
SMC_08		1	1					1
SMC_09		1	1					1
SMC_10		1	1					1
SMC_11		1	1					1
SMC_12		1	1					1

FOLIO REACTIVO	IDENTIFICADOR DE CUADERNILLO							
	PN2012SE C3ALF08	PN2012SE C3ALF09	PN2012SE C3ALF10	PN2012SE C3ALF11	PN2012SE C3ALF12	PN2012SE C3ALF13	PN2012SE C3ALF14	PN2012SE C3ALF15
SMD_01	1		1	1				
SMD_02	1		1	1				
SMD_03	1		1	1				
SMD_04	1		1	1				
SMD_05	1		1	1				
SMD_06	1		1	1				
SMD_07	1		1	1				
SMD_08	1		1	1				
SMD_09	1		1	1				
SMD_10	1		1	1				
SMD_11	1		1	1				
SMD_12	1		1	1				
SME_01		1		1	1			
SME_02		1		1	1			
SME_03		1		1	1			
SME_04		1		1	1			
SME_05		1		1	1			
SME_06		1		1	1			
SME_07		1		1	1			
SME_08		1		1	1			
SME_09		1		1	1			
SME_10		1		1	1			
SME_11		1		1	1			
SME_12		1		1	1			
SME_13		1		1	1			
SMF_01			1		1	1		
SMF_02			1		1	1		
SMF_03			1		1	1		
SMF_04			1		1	1		
SMF_05			1		1	1		
SMF_06			1		1	1		
SMF_07			1		1	1		
SMF_08			1		1	1		
SMF_09			1		1	1		
SMF_10			1		1	1		
SMF_11			1		1	1		
SMF_12			1		1	1		
SMF_13			1		1	1		
SMG_01				1		1	1	
SMG_02				1		1	1	

FOLIO REACTIVO	IDENTIFICADOR DE CUADERNILLO							
	PN2012SE C3ALF08	PN2012SE C3ALF09	PN2012SE C3ALF10	PN2012SE C3ALF11	PN2012SE C3ALF12	PN2012SE C3ALF13	PN2012SE C3ALF14	PN2012SE C3ALF15
SMG_03				1		1	1	
SMG_04				1		1	1	
SMG_05				1		1	1	
SMG_06				1		1	1	
SMG_07				1		1	1	
SMG_08				1		1	1	
SMG_09				1		1	1	
SMG_10				1		1	1	
SMG_11				1		1	1	
SMG_12				1		1	1	
SMH_01					1		1	1
SMH_02					1		1	1
SMH_03					1		1	1
SMH_04					1		1	1
SMH_05					1		1	1
SMH_06					1		1	1
SMH_07					1		1	1
SMH_08					1		1	1
SMH_09					1		1	1
SMH_10					1		1	1
SMH_11					1		1	1
SMH_12					1		1	1
SMH_13					1		1	1
<b>Total reactivos por cuadernillo</b>	37	38	37	37	39	37	38	37

**Fuente:** elaboración propia a partir de bases de datos proporcionadas al público por el INEE. Disponibles en: <http://www.inee.edu.mx/index.php/bases-de-datos/bases-de-datos-excale/excale-09-ciclo-2011-2012>



## Anexo 2. Script para el cálculo del DIF de la subescala SNPA mediante librería TAM de R

```
# Analisis DIF por SEXO para el subdominio SNPA
library("TAM")

#Directorio de trabajo.
setwd("C:/ejemplo")

#Leer los datos
SNPA_data <- read.csv("01a SNPA.csv")
SNPA_data

#Extraer la variable sexo
sexo <- SNPA_data[,42]
sexo

#Dividir en dos la base por variable sexo, Hombres=1 y Mujeres=2
scoredHom <- subset(SNPA_data[,1:40], sexo == "H")
scoredMuj <- subset(SNPA_data[,1:40], sexo == "M")
scoredHom
scoredMuj
summary(scoredHom)
summary(scoredMuj)

#Aplicación del Modelo de Diferencias
modHom <- tam.mml(scoredHom,constraint="items",irtmodel="PCM2")
modMuj <- tam.mml(scoredMuj,constraint="items",irtmodel="PCM2")

#Visualizar los resultados del modelo de diferencias
modHom$xi
modMuj$xi

###Análisis de la información para el Modelo de Diferencias
#Calcular los 40 reactivos de ambos sexos
itHo40<-(-sum(modHom$xi$xi))
itHo40
itHo40sd<-mean(modHom$xi$se.xi)
itHo40sd
itMu40<-(-sum(modMuj$xi$xi))
itMu40
itMu40sd<-mean(modMuj$xi$se.xi)
itMu40sd
deltaHo<-c(modHom$xi$xi,itHo40)
deltaHo
SEHo<-c(modHom$xi$se.xi,itHo40sd)
SEHo
deltaMu<-c(modMuj$xi$xi,itMu40)
deltaMu
SEMu<-c(modMuj$xi$se.xi,itMu40sd)
SEMu

#Información completa para los 40 reactivos
tablaSexo <- cbind(deltaHo,SEHo,deltaMu,SEMu)
tablaSexo
```

```

#Calcular DIF entre poblaciones
DIFSexo <-tablaSexo[,1]-tablaSexo[,3]
DIFSexo

#Diferencia estandarizada
zSexo <-DIFSexo/(sqrt((tablaSexo[,2])^2+(tablaSexo[,4])^2))
zSexo

#Categorías de DIF #A=ignorable #B=intermedio #C=grande
TamaqoSexo <- ifelse(abs(DIFSexo)<0.426,"A",(ifelse(abs(DIFSexo)>=0.638,"C","B")))
TamaqoSexo

#Grupo al que favorece el DIF
# si DIF y z son valores positivos se favorece al grupo focal (vulnerable)
FavoreceSexo <- ifelse(zSexo>0 , "Focal", "Referencia" )
FavoreceSexo

#Clasificación del DIF de acuerdo al grupo que favorece y su magnitud (ETS)
Clasificacion_ETSSexo <-ifelse(FavoreceSexo=="Referencia" & TamaqoSexo=="B","B-",
      (ifelse(FavoreceSexo=="Referencia" & TamaqoSexo=="C","C-",
        (ifelse(FavoreceSexo=="Focal" & TamaqoSexo=="B","B+",
          (ifelse(FavoreceSexo=="Focal" & TamaqoSexo=="C","C+", "**"))))))))
Clasificacion_ETSSexo

##Determinar si la diferencia es significativa
SigSexo <-ifelse((zSexo>1.96), "significativo",(ifelse(zSexo<(-1.96), "significativo", "**")))
SigSexo

#Resumen de la información
ResumenSexo <-
data.frame(cbind(tablaSexo,DIFSexo,zSexo,SigSexo,TamaqoSexo,FavoreceSexo,Clasificacion_ET
SSexo))
ResumenSexo

#Renombrando títulos de columnas #Sexo
names(ResumenSexo)[1]<-paste("Hom_dif")
names(ResumenSexo)[2]<-paste("Hom_SD")
names(ResumenSexo)[3]<-paste("Muj_dif")
names(ResumenSexo)[4]<-paste("Muj_SD")
names(ResumenSexo)[5]<-paste("Diferencia_grupos")
names(ResumenSexo)[6]<-paste("Indice_Rasch_DIF")
names(ResumenSexo)[7]<-paste("Significancia")
names(ResumenSexo)[8]<-paste("Tamaqo")
names(ResumenSexo)[9]<-paste("Favorece")
names(ResumenSexo)[10]<-paste("ETS")

#Exportar la tabla a Excel
write.csv(ResumenSexo, file="C:/ejemplo/01a SNPA Tabla TAM.csv")

```

### Anexo 3. Formato para Desglosar Procesos Cognitivos a Emplear en la Respuesta a Ítems del Excale-09 de Matemáticas

Responsable del desglose:
Fecha:

#### I. IDENTIFICACIÓN DEL REACTIVO-ESPECIFICACIÓN-CONTENIDO EVALUADO

Folio Reactivo:	Folio especificación:
a. Eje:	
b. Tema:	
c. Subtema:	
d. Conocimientos y habilidades:	
e. Especificación:	

II. Conocimientos y habilidades previos requeridos para contestar el reactivo correctamente:

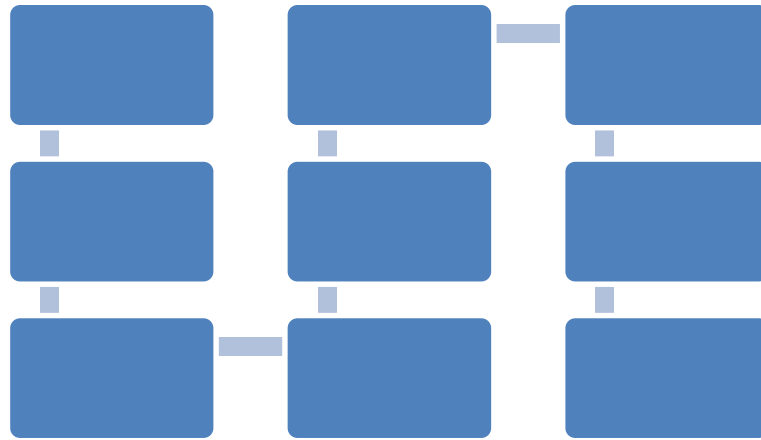
III. Actividades cognoscitivas involucradas en la contestación correcta del reactivo:

#### IV. Proceso cognitivo

Tomando en consideración las actividades cognoscitivas involucradas (apartado III) así como los conocimientos y habilidades requeridos por el estudiante para responder correctamente al reactivo (apartado II), definir el/los procesos cognitivos que se anticipa emplearán:

De manera especial, se requiere identificar el tipo de ruta cognitiva que tendrían que emplear, la cual podría ser: a) completamente lineal, b) lineal en un inicio y luego una bifurcación, c) en espiral, d) dos rutas en paralelo, u otra. Para ello, se pide que se complete alguno de los siguientes diagramas

a) Completamente lineal



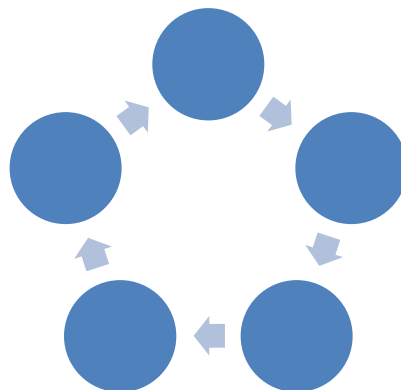
b) Lineal en un inicio y luego una bifurcación  
 - Parte común



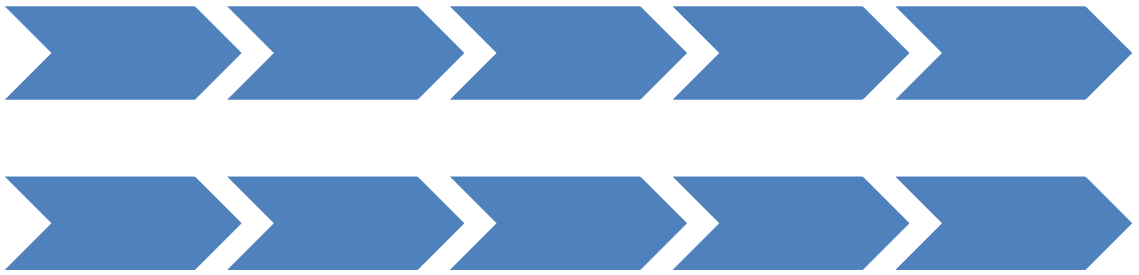
- A partir de la bifurcación



c) En espiral



d) Dos rutas en paralelo



e) Otro esquema

**V. Nivel cognitivo**

A partir de la revisión del proceso cognitivo que debe emplear el estudiante, definir el nivel cognitivo que demanda el reactivo. Para ello, es indispensable apoyarse en el texto *Niveles cognitivos establecidos en el Marco de Referencia del Excale-09 de Matemáticas*.

<b>Nivel cognitivo</b>	<b>Marcar con una X</b>
Conceptual	
Procedimental	
Resolución de problemas conocidos o típicos	
Resolución de problemas desconocidos o transferencia	
Nota: El nivel procedimental subsume el conceptual, y cualquiera de los dos niveles de resolución de problemas incluyen los niveles conceptual y/o procedimental.	

#### **Anexo 4. Niveles cognitivos establecidos en el Marco de Referencia del Excale-09 de Matemáticas**

A continuación se ofrece una explicación más amplia de los niveles cognitivos de los reactivos que conforman el Excale de matemáticas.

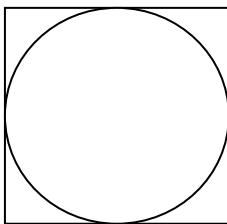
**Problemas de nivel conceptual:** se refiere a la comprensión de conceptos. Implica reconocer y generar ejemplos y contraejemplos de conceptos; generalizar características que se aplican a diferentes objetos que caen dentro de una categoría conceptual; reconocer, interpretar y aplicar símbolos y términos utilizados para representar conceptos; reconocer y generar definiciones de un concepto (no se trata de problemas que puedan resolverse con la simple evocación memorística). Ejemplos: identificar la altura en distintos tipos de triángulos, con distintas posiciones. Generar un ejemplo de triángulo rectángulo, identificar una bisectriz, etc.

**Nivel procedimental:** Se refiere a saber aplicar correctamente un procedimiento o algoritmo convencional. También puede implicar identificar en qué está mal un procedimiento presentado, o hacer modificaciones pertinentes a un procedimiento, a partir de condiciones especiales en un problema dado. Por ejemplo: resolver una ecuación cuadrática, ordenar los pasos para resolver un sistema de ecuaciones, identificar las instrucciones para realizar el trazo de una recta perpendicular a un segmento que pase por un punto dado fuera de él, etc.

**Nivel de resolución de problemas conocidos o típicos:** Se refiere a aplicar los conocimientos y habilidades en la resolución de una situación que puede ser modelada matemáticamente, pero cuyos datos se dan en términos coloquiales e informales. El problema debe corresponder a un modelo formal que presumiblemente sea conocido y haya sido practicado por el estudiante. Requiere que el estudiante elija los datos relevantes, la manera en que deben ser relacionados y aplique los procedimientos adecuados a esas relaciones para obtener o identificar el resultado correcto. Por ejemplo: Se va a cercar una parte de un terreno que colinda con un río y sólo se dispone de

material para construir 100 m lineales de barda con una altura de 1 metro. Si se quiere que la parte cercada tenga forma rectangular ¿cuáles serán las dimensiones del terreno de mayor área que se puede bardear?

**Nivel de resolución de problemas desconocidos o transferencia:** Es similar al nivel anterior, pero en este caso el problema debe corresponder a un modelo formal que presumiblemente sea desconocido y no haya sido practicado por el estudiante, pero del cual se considera que puede hacer una transferencia. Pretende conocer el grado en que el estudiante puede generalizar los principios que ha utilizado a resolver problemas conocidos a situaciones y esquemas lógicos diferentes a los que está habituado. Este nivel parece contradictorio con el principio general de las pruebas nacionales de evaluar solamente contenidos curriculares. Sin embargo, se incluye por dos razones: a) es importante tener una clasificación en la que se puedan incluir todos los casos, aunque a priori parezca que alguna de sus categorías no se utilizará; y, más importante, b) El uso de este nivel se justifica cuando se encuentre un contenido que está enunciado en los materiales curriculares de manera general pero no se encuentren ejemplos que lo aterricen en los materiales del alumno y el maestro. Ejemplo: El área del círculo es de  $6\text{m}^2$ . ¿Cuánto mide la diagonal del cuadrado que lo inscribe?





## **Anexo 5. Guía para la realización de los protocolos y reportes verbales mediante la técnica de pensamiento en voz alta y entrevista concurrente**

### **Paso 1. Revisar instalación del laboratorio cognitivo**

1. Mobiliario: 2 mesas de trabajo y 2 sillas; ubicadas de frente una con otra y a no más de 1.5 mts. de distancia entre ellas.
2. Equipo de videograbación: videocámara, trípode, micrófono lavalier, memoria SD, cargador eléctrico de la videocámara, extensión eléctrica.
3. Preparar el equipo de videograbación en su ubicación apropiada (evitando contraluz y que aparezcan en la toma tanto el entrevistado como el entrevistador), así como realizar todas las conexiones entre los diferentes componentes de la videograbación, de tal manera que esté lista para su utilización. Se empezará a videograbar pasos más adelante, pero desde ahora el equipo deberá estar preparado para funcionar.

### **Paso 2. Revisar materiales**

1. Material para entrevistado: cuadernillo con reactivos, lápiz, borrador, sacapuntas, consentimiento informado, hojas blancas
2. Material para entrevistador: guía de realización de entrevistas, formato de recolección de datos, ítems para realizar entrenamiento, lápiz y sacapuntas.

### **Paso 3. Bienvenida al entrevistado**

1. Breve presentación personal por parte del entrevistador
2. Agradecimiento al entrevistado por participar en el estudio
3. Breve presentación del estudio (sobre todo con aquellos participantes que no estuvieron durante la negociación de su participación)
4. Descripción al entrevistado de actividades que llevará a cabo durante la sesión

### **Paso 5. Entrenamiento para la técnica de pensamiento en voz alta**

1. Leerle al entrevistado el siguiente mensaje:  
*“Estamos interesados en conocer cómo respondes algunas preguntas de opción múltiple de un examen de matemáticas. Para ello, voy a mostrarte unos ítems y te pediré que pienses en voz alta cómo los vas solucionando, después te podría hacer algunas preguntas. Pensar en voz alta significa que quiero que me digas todo lo que pasa por tu mente desde el primer momento en que veas la pregunta del examen hasta el final cuando elijas la opción que consideres correcta. Mientras*

*revisas cada pregunta, es muy importante que te mantengas hablando, si en algún momento guardas silencio, te pediré que vuelvas a pensar en voz alta. Actúa como si estuvieras solo(a) en este salón y hablando para ti mismo(a). ¿Hasta ahora todo lo que te comenté es claro?”*

2. Se le ofrecerá al entrevistado un ejemplo de cómo pensar en voz alta. Para ello, el entrevistador podrá señalar cómo resuelve alguna operación (resta o multiplicación) de dos dígitos. Es importante que durante todo el proceso de respuesta, el entrevistador verbalice lo que piensa.

3. Primera práctica de pensamiento en voz alta. Resolviendo operaciones (no ítems todavía). Se le puede leer:

*Ahora vamos a practicar un poco, para que te acostumbres a pensar en voz alta. Te voy a pedir que esta hoja blanca que te acabo de proporcionar multipliques dos números. Así que piensa en voz alta mientras multiplicas 15 por 25.*

El entrevistado propiciará que el estudiante verbalice el proceso utilizado.

4. Ejemplo de cómo resolver un reactivo. Para que el estudiante se familiarice con la técnica de pensamiento en voz alta aplicada ítems, el investigador responderá un ítem a modo de ejemplo (que se proporcionará, junto con el resto de materiales) frente al participante para que pueda observar el comportamiento y desempeño que se espera de él. Pueden utilizarse las siguientes instrucciones:

*“Ahora voy a pensar en voz alta mientras resuelvo este problema. Eso significa que voy a decir todo lo que pasa por mi mente, incluyendo cómo entendí el problema que tengo que resolver, si hay palabras que no comprendo, si me resultan claras las indicaciones, si el problema presenta los datos suficientes para resolverlo, si recuerdo algún procedimiento que pueda utilizar para resolverlo, si en las opciones de respuesta viene alguna palabra que no comprendo o tengo dudas de lo que significa, etc.”*

En este punto, el entrevistador inicia a resolver el reactivo en voz alta.

Segunda práctica de pensamiento en voz alta. Se le solicita al entrevistado que responda el reactivo de ejemplo para el estudiante, recordándole que debe hablar cómo entiende el problema, si hay términos que no comprende, si utiliza alguna estrategia para descartar opciones de respuesta, etc.

El entrenamiento termina cuando el investigador observa que el participante presenta un nivel suficiente de desautomatización y de conciencia de los procesos y recursos cognitivos que usa para responder un ítem; de tal forma que en las verbalizaciones del participante puedan identificarse de forma clara los pasos que utilizó para responder el ítem y, de ser el caso, las dificultades que tuvo para su resolución.

5. Cierre del entrenamiento. Una vez concluido el entrenamiento, se le recuerda al entrevistado que el principal objetivo por el que se le está entrevistando es conocer el proceso que empleó para responder cada reactivo. Al final, se le pregunta si tiene alguna duda, se le resuelve y, ante la ausencia de dudas, se continúa con el siguiente paso.

Paso 6. Aplicación de técnica de pensamiento en voz alta y entrevista concurrente.

2. Iniciar la videograbación, cerciorándose que el micrófono lavalier está funcionando.
3. Entregar el cuadernillo al estudiante y pedirle que lea las indicaciones generales, donde se le hará saber que en cada página sólo encontrará un ítem, y que no podrá cambiar de página hasta que el investigador se lo haga saber.
4. Para cada reactivo se realizará el siguiente ciclo. Es decir, se repetirá 18 veces:
  - a. Se le pedirá al estudiante que verbalice sus pensamientos. Durante este tiempo el investigador tendrá principalmente dos funciones: i) asegurarse que el entrevistado piense en voz alta mientras responde el ítem e, ii) identificar lagunas de información para registrar en el *Formato de recolección de datos* las preguntas adicionales que formulará posteriormente para subsanar dichos vacíos de información.
  - b. Una vez que el estudiante concluya de responder un reactivo en específico, se realizará el sondeo a partir de las preguntas eje que están incluidas en el *Formato de recolección de datos* y de las preguntas adicionales que el investigador fue redactando a partir de la técnica de pensamiento en voz alta; todo ello con el objetivo de detallar el proceso empleado y, en especial, identificar ciertos términos que por la forma en que se comprenden, puedan alterar el resultado del reactivo.

Paso 7. Agradecimiento al entrevistado

Paso 8. Integración del expediente de cada entrevistado. El cual deberá contener:

1. *Formato de recolección de datos* debidamente complementado.
2. Archivo con la videograbación. El cual deberá nombrarse con el mismo número de Folio del entrevistado.
3. El cuadernillo del examen respondido por el entrevistado.
4. Consentimiento informado, firmado por el entrevistado.

## Anexo 6. Formato de recolección de datos

Fecha de aplicación: \_\_\_ de \_\_\_\_\_ de 2016

<b>Datos del participante</b>	
Folio del entrevistado: _____	
Nombre: _____	Edad: _____
Correo electrónico: _____	
Escuela: _____	Grado escolar: _____
Promedio de los dos últimos bimestres en Matemáticas (3°): _____	

## ÍTEM X<sup>10</sup>

### 1. Pie del reactivo

Palabras leídas incorrectamente:

Términos en los que el entrevistado verbalizó confusión o desconocimiento de su significado:

Preguntas eje para sondeo concurrente:

- *¿Qué tan clara te parece la pregunta X?*
- *¿Hubo alguna palabra o concepto que no conocías? ¿o hubo alguna palabra que te causara confusión sobre su significado?*
- *¿Te resulta claro lo que tienes que hacer para resolver el problema que se plantea?*
- *En la pregunta, ¿hubo alguna situación con la que te sintieras incómodo o agredido?*

Preguntas adicionales formuladas por entrevistador:

### 2. Proceso de respuesta

Palabras leídas de manera incorrecta en las opciones de respuesta:

Identificación de elementos del contexto que obstaculizaron o favorecieron que eligiera la opción de respuesta marcada:

¿El entrevistado respondió correctamente el ítem?

- Sí, y utilizó un procedimiento de solución de manera correcta
- Sí, aunque hubo algunos elementos del proceso de solución que desconocía
- Sí, aunque aparentemente adivinó
- No, aunque el procedimiento que empleó fue el correcto
- No, y desconocía el procedimiento correcto que debía utilizar

Preguntas eje:

- *Explícame, paso a paso, el procedimiento que utilizaste para responder la pregunta*
- *¿Cuándo respondiste a la pregunta X, la respuesta que tú pensaste fue distinta a las opciones de respuesta que venían escritas? ¿Cuál era tu respuesta? ¿Por qué llegaste a esa solución?*
- *En las opciones de respuesta, ¿hubo alguna palabra que no entendiste o que te confundiera? ¿Cuál?*
- *Desde tu punto de vista, ¿hay una sola opción de respuesta que es la correcta?*
- *Alguna de las opciones que no elegiste, ¿te parece que también podría ser la respuesta correcta?*

Preguntas adicionales formuladas por entrevistador:

<sup>10</sup> El formato es igual para cada uno de los 18 ítems, por eso sólo se presenta una sola vez.

## Anexo 7. Carta de Consentimiento

Mediante la presente reconozco participar en el estudio titulado *Obtención de evidencias de validez con enfoque de equidad de los Excale de Español y Matemáticas para tercero de secundaria en México.*

Mi participación consistirá en responder 18 reactivos de la asignatura de Matemáticas que fueron parte de un examen nacional que aplicó en 2012 el Instituto Nacional para la Evaluación de la Educación a alumnos de tercero de secundaria. Asimismo, mi participación consistirá en responder las preguntas que sobre esos 18 reactivos me hará el investigador a cargo de realizar la entrevista.

Los resultados de esta investigación, a partir de mi participación y la de otros profesores y alumnos, servirán para evaluar la calidad de los reactivos y hacer recomendaciones para su mejora. Dado que el estudio se utilizará sólo con fines académicos (no comerciales), autorizo para que mi participación sea filmada, con el compromiso del equipo de investigadores de que se mantendrá el anonimato y confidencialidad de la información; lo cual implica que de ninguna manera se podrán dar a conocer los resultados que en lo individual yo haya obtenido, sino sólo de forma agrupada y con los mecanismos necesarios para preservar mi anonimato.

Porque sólo responderé una sección del examen de Matemáticas, estoy consciente que no se me podrá otorgar una calificación.

	Por parte del equipo de investigación	Por parte del alumno	Por parte del padre de familia
Nombre:			
Firma:			
Fecha:			

**Anexo 8. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala *Sentido Numérico y Pensamiento Algebraico* del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_01	H	0.77	M	0.77	0
SMA_02	H	-1.26	M	-0.89	-0.36
SMA_05	H	-0.05	M	0.01	-0.06
SMA_08	H	-1.74	M	-1.74	0
SMA_12	H	-0.06	M	-0.13	0.08
SMB_01	H	-0.01	M	0.38	-0.38
SMB_02	H	0.39	M	0.39	0
SMB_03	H	-0.09	M	-0.09	0
SMB_04	H	0.91	M	0.86	0.05
SMB_05	H	-1.45	M	-1.49	0.04
SMC_01	H	-0.08	M	-0.06	-0.02
SMC_02	H	-0.33	M	-0.47	0.13
SMC_03	H	0.36	M	0.5	-0.14
SMC_05	H	1.6	M	1.77	-0.17
SMC_06	H	0.04	M	0.02	0.02
SMD_01	H	0.94	M	0.84	0.1
SMD_10	H	-0.43	M	-0.85	0.42
SMD_12	H	0.64	M	0.82	-0.18
SME_01	H	0.11	M	0.17	-0.06
SME_02	H	-0.43	M	-0.61	0.17
SME_03	H	0.2	M	-0.06	0.26
SMF_01	H	-0.22	M	-0.36	0.15
SMF_02	H	-1.31	M	-1.62	0.31
SMF_03	H	0.51	M	0.66	-0.15
SMF_04	H	-0.18	M	-0.15	-0.02
SMF_05	H	-0.06	M	-0.15	0.09
SMG_01	H	1.19	M	1.1	0.08
SMG_02	H	-0.03	M	-0.12	0.09
SMG_03	H	0.47	M	0.36	0.11
SMG_04	H	-1.05	M	-0.64	-0.41
SMG_05	H	-0.04	M	0.07	-0.1
SMG_08	H	0.1	M	0.19	-0.1
SMG_09	H	-0.85	M	-0.66	-0.19
SMG_10	H	-0.18	M	-0.18	0
SMH_01	H	-0.74	M	-0.84	0.1
SMH_02	H	0.4	M	0.45	-0.05
SMH_03	H	0.03	M	0.03	0

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMH_04	H	1.32	M	1.39	-0.07
SMH_05	H	-0.14	M	-0.32	0.18
SMH_07	H	0.72	M	0.72	0

\* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.



**Anexo 9. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_03	H	0.13	M	0.28	-0.15
SMA_04	H	-0.2	M	0.03	-0.22
SMA_06	H	0.03	M	-0.11	0.13
SMA_07	H	0.06	M	-0.31	0.36
SMB_06	H	-0.23	M	-0.01	-0.22
SMB_07	H	-1.18	M	-1.37	0.19
SMB_08	H	0.75	M	0.48	0.27
SMB_09	H	-0.46	M	-0.39	-0.06
SMB_10	H	0.24	M	0.3	-0.06
SMD_02	H	1.48	M	1.48	0
SMD_03	H	-0.94	M	-1	0.06
SMD_04	H	0.75	M	0.63	0.12
SMD_05	H	-0.34	M	-0.43	0.09
SMD_06	H	0.86	M	1.04	-0.18
SMD_07	H	0.89	M	0.97	-0.07
SMD_08	H	-1.53	M	-1.6	0.07
SMD_09	H	-1.06	M	-0.86	-0.19
SME_04	H	0.92	M	0.9	0.02
SME_05	H	-0.6	M	-0.53	-0.07
SME_06	H	-1.01	M	-0.85	-0.16
SME_07	H	0.28	M	0.28	0
SME_08	H	-0.07	M	-0.07	0
SMF_06	H	0.22	M	0.13	0.09
SMF_07	H	-0.25	M	-0.48	0.23
SMF_08	H	1.22	M	1.32	-0.1
SMF_10	H	-1.23	M	-1.05	-0.18
SMG_06	H	-1.03	M	-1.19	0.16
SMG_07	H	2.48	M	2.48	0
SMH_08	H	0.67	M	0.67	0
SMH_09	H	-1.01	M	-0.62	-0.39
SMH_10	H	-0.01	M	-0.01	0
SMH_11	H	0.16	M	-0.05	0.21
* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$ & $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).					
<b>Fuente:</b> elaboración propia.					

**Anexo 10. Tamaño del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_09	H	1.55	M	1.46	0.09
SMA_10	H	0.35	M	0.18	0.16
SMA_11	H	1.08	M	1.08	0
SMB_11	H	-0.87	M	-0.52	-0.34
SMB_12	H	1.05	M	1.18	-0.13
SMB_13	H	1.29	M	1.21	0.08
SMC_04	H	-0.1	M	-0.2	0.1
SMC_07	H	0.42	M	0.59	-0.17
SMC_08	H	-2.38	M	-2.74	0.36
SMC_09	H	-0.12	M	-0.12	0
SMC_10	H	0.65	M	0.91	-0.26
SMC_11	H	-0.4	M	-0.4	0
SMC_12	H	-0.04	M	-0.24	0.2
SMD_11	H	0.19	M	0.19	0
SME_09	H	0.72	M	0.85	-0.13
SME_10	H	-1.23	M	-0.93	-0.3
SME_11	H	0.59	M	0.59	0
SME_12	H	-0.87	M	-0.87	0
SME_13	H	0.42	M	0.19	0.23
SMF_09	H	-1.1	M	-0.75	-0.35
SMF_11	H	0.07	M	-0.14	0.21
SMF_12	H	0.04	M	0.1	-0.06
SMF_13	H	1.14	M	1.28	-0.13
SMG_11	H	0.23	M	0.11	0.12
SMG_12	H	-1.52	M	-1.41	-0.11
SMH_06	H	0.31	M	-0.11	0.42
SMH_12	H	-0.54	M	-0.39	-0.15
SMH_13	H	-0.9	M	-1.1	0.2

\* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 11. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_01	Hom-Tercil 1	0.72	Muj-Tercil 1	0.71	0.01
SMA_02	Hom-Tercil 1	-0.9	Muj-Tercil 1	-0.41	-0.49
SMA_05	Hom-Tercil 1	-0.14	Muj-Tercil 1	-0.12	-0.02
SMA_08	Hom-Tercil 1	-1.52	Muj-Tercil 1	-1.49	-0.03
SMA_12	Hom-Tercil 1	0.1	Muj-Tercil 1	0.03	0.07
SMB_01	Hom-Tercil 1	0.38	Muj-Tercil 1	0.59	-0.21
SMB_02	Hom-Tercil 1	0.26	Muj-Tercil 1	0.62	-0.35
SMB_03	Hom-Tercil 1	-0.54	Muj-Tercil 1	-0.46	-0.09
SMB_04	Hom-Tercil 1	0.29	Muj-Tercil 1	0.66	-0.37
SMB_05	Hom-Tercil 1	-1.2	Muj-Tercil 1	-1.19	0
SMC_01	Hom-Tercil 1	-0.22	Muj-Tercil 1	-0.21	-0.01
SMC_02	Hom-Tercil 1	-0.46	Muj-Tercil 1	-0.46	-0.01
SMC_03	Hom-Tercil 1	0.38	Muj-Tercil 1	0.25	0.13
SMC_05	Hom-Tercil 1	1.07	Muj-Tercil 1	1.46	-0.39
SMC_06	Hom-Tercil 1	0.41	Muj-Tercil 1	0.33	0.07
SMD_01	Hom-Tercil 1	0.56	Muj-Tercil 1	0.53	0.03
SMD_10	Hom-Tercil 1	-0.21	Muj-Tercil 1	-0.69	0.48
SMD_12	Hom-Tercil 1	0.62	Muj-Tercil 1	1.08	-0.46
SME_01	Hom-Tercil 1	0.23	Muj-Tercil 1	0.38	-0.15
SME_02	Hom-Tercil 1	-0.81	Muj-Tercil 1	-1.06	0.25
SME_03	Hom-Tercil 1	0.12	Muj-Tercil 1	-0.22	0.34
SMF_01	Hom-Tercil 1	-0.45	Muj-Tercil 1	-0.55	0.1
SMF_02	Hom-Tercil 1	-0.96	Muj-Tercil 1	-1.46	0.5
SMF_03	Hom-Tercil 1	0.53	Muj-Tercil 1	0.62	-0.09
SMF_04	Hom-Tercil 1	-0.42	Muj-Tercil 1	-0.33	-0.08
SMF_05	Hom-Tercil 1	-0.23	Muj-Tercil 1	-0.05	-0.18
SMG_01	Hom-Tercil 1	0.62	Muj-Tercil 1	0.66	-0.04
SMG_02	Hom-Tercil 1	-0.04	Muj-Tercil 1	-0.15	0.11
SMG_03	Hom-Tercil 1	0.23	Muj-Tercil 1	0.17	0.06
SMG_04	Hom-Tercil 1	-0.94	Muj-Tercil 1	-0.57	-0.37
SMG_05	Hom-Tercil 1	0.07	Muj-Tercil 1	0.54	-0.47
SMG_08	Hom-Tercil 1	0.08	Muj-Tercil 1	0.25	-0.17
SMG_09	Hom-Tercil 1	-0.81	Muj-Tercil 1	-0.64	-0.17
SMG_10	Hom-Tercil 1	0.12	Muj-Tercil 1	0.01	0.12
SMH_01	Hom-Tercil 1	-0.84	Muj-Tercil 1	-1.06	0.22
SMH_02	Hom-Tercil 1	0.57	Muj-Tercil 1	0.47	0.1
SMH_03	Hom-Tercil 1	-0.03	Muj-Tercil 1	-0.31	0.28
SMH_04	Hom-Tercil 1	0.63	Muj-Tercil 1	0.98	-0.36
SMH_05	Hom-Tercil 1	0.14	Muj-Tercil 1	-0.19	0.33

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMH_07	Hom-Tercil 1	0.62	Muj-Tercil 1	0.29	0.32
SMA_01	Hom-Tercil 2	0.87	Muj-Tercil 2	0.77	0.1
SMA_02	Hom-Tercil 2	-1.38	Muj-Tercil 2	-1.04	-0.34
SMA_05	Hom-Tercil 2	-0.02	Muj-Tercil 2	0.06	-0.07
SMA_08	Hom-Tercil 2	-1.86	Muj-Tercil 2	-1.88	0.03
SMA_12	Hom-Tercil 2	-0.05	Muj-Tercil 2	-0.16	0.1
SMB_01	Hom-Tercil 2	-0.03	Muj-Tercil 2	0.4	-0.43
SMB_02	Hom-Tercil 2	0.39	Muj-Tercil 2	0.39	0
SMB_03	Hom-Tercil 2	-0.14	Muj-Tercil 2	-0.09	-0.05
SMB_04	Hom-Tercil 2	0.93	Muj-Tercil 2	0.82	0.11
SMB_05	Hom-Tercil 2	-1.52	Muj-Tercil 2	-1.61	0.09
SMC_01	Hom-Tercil 2	-0.08	Muj-Tercil 2	0	-0.08
SMC_02	Hom-Tercil 2	-0.32	Muj-Tercil 2	-0.49	0.17
SMC_03	Hom-Tercil 2	0.35	Muj-Tercil 2	0.55	-0.21
SMC_05	Hom-Tercil 2	1.65	Muj-Tercil 2	1.8	-0.15
SMC_06	Hom-Tercil 2	0.02	Muj-Tercil 2	-0.06	0.08
SMD_01	Hom-Tercil 2	0.94	Muj-Tercil 2	0.84	0.1
SMD_10	Hom-Tercil 2	-0.47	Muj-Tercil 2	-0.89	0.42
SMD_12	Hom-Tercil 2	0.68	Muj-Tercil 2	0.87	-0.18
SME_01	Hom-Tercil 2	0.11	Muj-Tercil 2	0.16	-0.06
SME_02	Hom-Tercil 2	-0.37	Muj-Tercil 2	-0.52	0.15
SME_03	Hom-Tercil 2	0.16	Muj-Tercil 2	-0.07	0.23
SMF_01	Hom-Tercil 2	-0.19	Muj-Tercil 2	-0.34	0.15
SMF_02	Hom-Tercil 2	-1.46	Muj-Tercil 2	-1.68	0.22
SMF_03	Hom-Tercil 2	0.5	Muj-Tercil 2	0.67	-0.16
SMF_04	Hom-Tercil 2	-0.11	Muj-Tercil 2	-0.08	-0.03
SMF_05	Hom-Tercil 2	-0.02	Muj-Tercil 2	-0.17	0.15
SMG_01	Hom-Tercil 2	1.15	Muj-Tercil 2	1.06	0.09
SMG_02	Hom-Tercil 2	0.01	Muj-Tercil 2	-0.11	0.12
SMG_03	Hom-Tercil 2	0.52	Muj-Tercil 2	0.38	0.14
SMG_04	Hom-Tercil 2	-1.07	Muj-Tercil 2	-0.64	-0.43
SMG_05	Hom-Tercil 2	-0.03	Muj-Tercil 2	0.02	-0.05
SMG_08	Hom-Tercil 2	0.12	Muj-Tercil 2	0.19	-0.07
SMG_09	Hom-Tercil 2	-0.85	Muj-Tercil 2	-0.65	-0.2
SMG_10	Hom-Tercil 2	-0.2	Muj-Tercil 2	-0.15	-0.05
SMH_01	Hom-Tercil 2	-0.7	Muj-Tercil 2	-0.79	0.09
SMH_02	Hom-Tercil 2	0.35	Muj-Tercil 2	0.42	-0.07
SMH_03	Hom-Tercil 2	0.03	Muj-Tercil 2	0.11	-0.08
SMH_04	Hom-Tercil 2	1.35	Muj-Tercil 2	1.41	-0.06
SMH_05	Hom-Tercil 2	-0.17	Muj-Tercil 2	-0.35	0.18
SMH_07	Hom-Tercil 2	0.82	Muj-Tercil 2	0.8	0.02
SMA_01	Hom-Tercil 3	0.07	Muj-Tercil 3	0.53	-0.45
SMA_02	Hom-Tercil 3	-2.19	Muj-Tercil 3	-2.14	-0.05

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_05	Hom-Tercil 3	-0.32	Muj-Tercil 3	-0.16	-0.15
SMA_08	Hom-Tercil 3	-2.73	Muj-Tercil 3	-1.8	-0.93
SMA_12	Hom-Tercil 3	-0.72	Muj-Tercil 3	-0.62	-0.1
SMB_01	Hom-Tercil 3	-1.02	Muj-Tercil 3	-1.09	0.07
SMB_02	Hom-Tercil 3	0.81	Muj-Tercil 3	-0.23	1.04
SMB_03	Hom-Tercil 3	1.82	Muj-Tercil 3	1.38	0.44
SMB_04	Hom-Tercil 3	1.55	Muj-Tercil 3	1.58	-0.03
SMB_05	Hom-Tercil 3	-7.85>	Muj-Tercil 3	-3.6	-4.25
SMC_01	Hom-Tercil 3	-0.12	Muj-Tercil 3	-0.96	0.83
SMC_02	Hom-Tercil 3	0.32	Muj-Tercil 3	0.54	-0.22
SMC_03	Hom-Tercil 3	0.47	Muj-Tercil 3	0.56	-0.1
SMC_05	Hom-Tercil 3	1.84	Muj-Tercil 3	1.98	-0.14
SMC_06	Hom-Tercil 3	-1.23	Muj-Tercil 3	-0.55	-0.68
SMD_01	Hom-Tercil 3	1.4	Muj-Tercil 3	1.24	0.16
SMD_10	Hom-Tercil 3	-0.87	Muj-Tercil 3	-1.29	0.42
SMD_12	Hom-Tercil 3	0.16	Muj-Tercil 3	-0.02	0.18
SME_01	Hom-Tercil 3	-0.39	Muj-Tercil 3	-0.66	0.27
SME_02	Hom-Tercil 3	0.41	Muj-Tercil 3	0.35	0.06
SME_03	Hom-Tercil 3	1.22	Muj-Tercil 3	0.82	0.4
SMF_01	Hom-Tercil 3	0.36	Muj-Tercil 3	0.29	0.07
SMF_02	Hom-Tercil 3	-1.85	Muj-Tercil 3	-3.12	1.27
SMF_03	Hom-Tercil 3	0.59	Muj-Tercil 3	0.59	0
SMF_04	Hom-Tercil 3	-0.51	Muj-Tercil 3	-0.66	0.14
SMF_05	Hom-Tercil 3	-0.03	Muj-Tercil 3	-0.14	0.1
SMG_01	Hom-Tercil 3	1.99	Muj-Tercil 3	2.06	-0.07
SMG_02	Hom-Tercil 3	-0.87	Muj-Tercil 3	-0.25	-0.62
SMG_03	Hom-Tercil 3	0.36	Muj-Tercil 3	0.6	-0.23
SMG_04	Hom-Tercil 3	-1.67	Muj-Tercil 3	-1.66	-0.01
SMG_05	Hom-Tercil 3	-0.46	Muj-Tercil 3	-0.21	-0.25
SMG_08	Hom-Tercil 3	-0.31	Muj-Tercil 3	-0.02	-0.29
SMG_09	Hom-Tercil 3	-1.24	Muj-Tercil 3	-1.09	-0.15
SMG_10	Hom-Tercil 3	-1.73	Muj-Tercil 3	-2.1	0.36
SMH_01	Hom-Tercil 3	-1.15	Muj-Tercil 3	-1.58	0.44
SMH_02	Hom-Tercil 3	0.86	Muj-Tercil 3	1.41	-0.55
SMH_03	Hom-Tercil 3	-0.17	Muj-Tercil 3	0.03	-0.2
SMH_04	Hom-Tercil 3	1.9	Muj-Tercil 3	1.7	0.2
SMH_05	Hom-Tercil 3	-1.05	Muj-Tercil 3	-0.43	-0.63
SMH_07	Hom-Tercil 3	-0.57	Muj-Tercil 3	-0.33	-0.24

\* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 12. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_03	Hom-Tercil 1	0.34	Muj-Tercil 1	0.26	0.09
SMA_04	Hom-Tercil 1	0.24	Muj-Tercil 1	0.28	-0.04
SMA_06	Hom-Tercil 1	-0.08	Muj-Tercil 1	-0.33	0.25
SMA_07	Hom-Tercil 1	0.07	Muj-Tercil 1	-0.31	0.39
SMB_06	Hom-Tercil 1	-0.46	Muj-Tercil 1	-0.11	-0.34
SMB_07	Hom-Tercil 1	-1.05	Muj-Tercil 1	-1.28	0.23
SMB_08	Hom-Tercil 1	0.49	Muj-Tercil 1	0.24	0.25
SMB_09	Hom-Tercil 1	-0.29	Muj-Tercil 1	-0.05	-0.23
SMB_10	Hom-Tercil 1	0.18	Muj-Tercil 1	0.11	0.06
SMD_02	Hom-Tercil 1	1.5	Muj-Tercil 1	1.61	-0.1
SMD_03	Hom-Tercil 1	-0.69	Muj-Tercil 1	-0.87	0.18
SMD_04	Hom-Tercil 1	0.15	Muj-Tercil 1	0	0.15
SMD_05	Hom-Tercil 1	-0.66	Muj-Tercil 1	-0.84	0.18
SMD_06	Hom-Tercil 1	1.15	Muj-Tercil 1	1.05	0.1
SMD_07	Hom-Tercil 1	0.58	Muj-Tercil 1	0.93	-0.36
SMD_08	Hom-Tercil 1	-1.31	Muj-Tercil 1	-1.38	0.07
SMD_09	Hom-Tercil 1	-0.81	Muj-Tercil 1	-0.45	-0.36
SME_04	Hom-Tercil 1	0.37	Muj-Tercil 1	0.64	-0.27
SME_05	Hom-Tercil 1	-0.61	Muj-Tercil 1	-0.64	0.03
SME_06	Hom-Tercil 1	-0.55	Muj-Tercil 1	-0.56	0.01
SME_07	Hom-Tercil 1	0.33	Muj-Tercil 1	0.28	0.05
SME_08	Hom-Tercil 1	-0.35	Muj-Tercil 1	-0.27	-0.08
SMF_06	Hom-Tercil 1	-0.08	Muj-Tercil 1	0	-0.08
SMF_07	Hom-Tercil 1	-0.53	Muj-Tercil 1	-0.51	-0.02
SMF_08	Hom-Tercil 1	1.09	Muj-Tercil 1	0.98	0.11
SMF_10	Hom-Tercil 1	-1.07	Muj-Tercil 1	-0.78	-0.29
SMG_06	Hom-Tercil 1	-0.83	Muj-Tercil 1	-0.94	0.1
SMG_07	Hom-Tercil 1	3.01	Muj-Tercil 1	2.79	0.22
SMH_08	Hom-Tercil 1	0.42	Muj-Tercil 1	0.38	0.04
SMH_09	Hom-Tercil 1	-0.94	Muj-Tercil 1	-0.53	-0.42
SMH_10	Hom-Tercil 1	0.29	Muj-Tercil 1	0.07	0.22
SMH_11	Hom-Tercil 1	-0.51	Muj-Tercil 1	-0.48	-0.03
SMA_03	Hom-Tercil 2	0.08	Muj-Tercil 2	0.3	-0.22
SMA_04	Hom-Tercil 2	-0.31	Muj-Tercil 2	-0.02	-0.29
SMA_06	Hom-Tercil 2	0.04	Muj-Tercil 2	-0.06	0.11
SMA_07	Hom-Tercil 2	0.05	Muj-Tercil 2	-0.3	0.35
SMB_06	Hom-Tercil 2	-0.11	Muj-Tercil 2	0.04	-0.16

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMB_07	Hom-Tercil 2	-1.28	Muj-Tercil 2	-1.43	0.15
SMB_08	Hom-Tercil 2	0.81	Muj-Tercil 2	0.51	0.3
SMB_09	Hom-Tercil 2	-0.52	Muj-Tercil 2	-0.51	-0.01
SMB_10	Hom-Tercil 2	0.27	Muj-Tercil 2	0.37	-0.1
SMD_02	Hom-Tercil 2	1.52	Muj-Tercil 2	1.51	0
SMD_03	Hom-Tercil 2	-1.04	Muj-Tercil 2	-1.05	0.02
SMD_04	Hom-Tercil 2	0.79	Muj-Tercil 2	0.69	0.1
SMD_05	Hom-Tercil 2	-0.23	Muj-Tercil 2	-0.3	0.07
SMD_06	Hom-Tercil 2	0.86	Muj-Tercil 2	1.09	-0.22
SMD_07	Hom-Tercil 2	0.98	Muj-Tercil 2	0.98	0
SMD_08	Hom-Tercil 2	-1.67	Muj-Tercil 2	-1.73	0.06
SMD_09	Hom-Tercil 2	-1.16	Muj-Tercil 2	-1	-0.16
SME_04	Hom-Tercil 2	0.98	Muj-Tercil 2	0.9	0.08
SME_05	Hom-Tercil 2	-0.59	Muj-Tercil 2	-0.48	-0.11
SME_06	Hom-Tercil 2	-1.19	Muj-Tercil 2	-0.95	-0.24
SME_07	Hom-Tercil 2	0.28	Muj-Tercil 2	0.25	0.03
SME_08	Hom-Tercil 2	0.01	Muj-Tercil 2	-0.03	0.04
SMF_06	Hom-Tercil 2	0.25	Muj-Tercil 2	0.13	0.12
SMF_07	Hom-Tercil 2	-0.21	Muj-Tercil 2	-0.49	0.28
SMF_08	Hom-Tercil 2	1.25	Muj-Tercil 2	1.39	-0.14
SMF_10	Hom-Tercil 2	-1.28	Muj-Tercil 2	-1.14	-0.15
SMG_06	Hom-Tercil 2	-1.12	Muj-Tercil 2	-1.3	0.18
SMG_07	Hom-Tercil 2	2.64	Muj-Tercil 2	2.7	-0.07
SMH_08	Hom-Tercil 2	0.7	Muj-Tercil 2	0.67	0.03
SMH_09	Hom-Tercil 2	-1.03	Muj-Tercil 2	-0.64	-0.39
SMH_10	Hom-Tercil 2	-0.07	Muj-Tercil 2	-0.01	-0.06
SMH_11	Hom-Tercil 2	0.3	Muj-Tercil 2	0.03	0.28
SMA_03	Hom-Tercil 3	-0.23	Muj-Tercil 3	-0.49	0.25
SMA_04	Hom-Tercil 3	-1.76	Muj-Tercil 3	-1.17	-0.59
SMA_06	Hom-Tercil 3	0.54	Muj-Tercil 3	0.82	-0.27
SMA_07	Hom-Tercil 3	0.24	Muj-Tercil 3	-0.45	0.69
SMB_06	Hom-Tercil 3	-0.85	Muj-Tercil 3	-0.21	-0.64
SMB_07	Hom-Tercil 3	-1.71	Muj-Tercil 3	-1.57	-0.14
SMB_08	Hom-Tercil 3	0.9	Muj-Tercil 3	1.19	-0.29
SMB_09	Hom-Tercil 3	-0.96	Muj-Tercil 3	-1.39	0.44
SMB_10	Hom-Tercil 3	0.12	Muj-Tercil 3	0	0.12
SMD_02	Hom-Tercil 3	0.92	Muj-Tercil 3	0.87	0.04
SMD_03	Hom-Tercil 3	-1.85	Muj-Tercil 3	-0.73	-1.12
SMD_04	Hom-Tercil 3	2.3	Muj-Tercil 3	2.08	0.22
SMD_05	Hom-Tercil 3	-0.28	Muj-Tercil 3	-0.31	0.03
SMD_06	Hom-Tercil 3	-0.19	Muj-Tercil 3	0.01	-0.2

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMD_07	Hom-Tercil 3	0.51	Muj-Tercil 3	0.69	-0.18
SMD_08	Hom-Tercil 3	-2.75	Muj-Tercil 3	-2.08	-0.67
SMD_09	Hom-Tercil 3	-2.19	Muj-Tercil 3	-1.91	-0.29
SME_04	Hom-Tercil 3	1.66	Muj-Tercil 3	1.31	0.35
SME_05	Hom-Tercil 3	-0.74	Muj-Tercil 3	-1.19	0.46
SME_06	Hom-Tercil 3	-1.54	Muj-Tercil 3	-2.08	0.54
SME_07	Hom-Tercil 3	0.45	Muj-Tercil 3	0.43	0.02
SME_08	Hom-Tercil 3	0.23	Muj-Tercil 3	0.5	-0.27
SMF_06	Hom-Tercil 3	0.99	Muj-Tercil 3	0.71	0.28
SMF_07	Hom-Tercil 3	0.8	Muj-Tercil 3	0.33	0.47
SMF_08	Hom-Tercil 3	1.04	Muj-Tercil 3	0.91	0.13
SMF_10	Hom-Tercil 3	-1.5	Muj-Tercil 3	-2.12	0.62
SMG_06	Hom-Tercil 3	-1.94	Muj-Tercil 3	-1.59	-0.35
SMG_07	Hom-Tercil 3	1.46	Muj-Tercil 3	1.27	0.19
SMH_08	Hom-Tercil 3	1.21	Muj-Tercil 3	1.77	-0.56
SMH_09	Hom-Tercil 3	-1.43	Muj-Tercil 3	-1.14	-0.29
SMH_10	Hom-Tercil 3	-0.77	Muj-Tercil 3	-1.27	0.51
SMH_11	Hom-Tercil 3	1.54	Muj-Tercil 3	1.52	0.02

\* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.



**Anexo 13. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Sexo**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_09	Hom-Tercil 1	1.28	Muj-Tercil 1	1.34	-0.06
SMA_10	Hom-Tercil 1	0.18	Muj-Tercil 1	-0.11	0.3
SMA_11	Hom-Tercil 1	0.59	Muj-Tercil 1	0.85	-0.26
SMB_11	Hom-Tercil 1	-0.59	Muj-Tercil 1	-0.35	-0.24
SMB_12	Hom-Tercil 1	0.45	Muj-Tercil 1	0.62	-0.16
SMB_13	Hom-Tercil 1	0.88	Muj-Tercil 1	0.94	-0.06
SMC_04	Hom-Tercil 1	0.12	Muj-Tercil 1	-0.09	0.21
SMC_07	Hom-Tercil 1	0.35	Muj-Tercil 1	0.5	-0.15
SMC_08	Hom-Tercil 1	-1.94	Muj-Tercil 1	-2.44	0.5
SMC_09	Hom-Tercil 1	0.18	Muj-Tercil 1	0.3	-0.12
SMC_10	Hom-Tercil 1	-0.27	Muj-Tercil 1	0.04	-0.31
SMC_11	Hom-Tercil 1	-0.05	Muj-Tercil 1	-0.14	0.09
SMC_12	Hom-Tercil 1	-0.14	Muj-Tercil 1	-0.46	0.32
SMD_11	Hom-Tercil 1	-0.2	Muj-Tercil 1	-0.28	0.08
SME_09	Hom-Tercil 1	0.03	Muj-Tercil 1	0.23	-0.2
SME_10	Hom-Tercil 1	-1.1	Muj-Tercil 1	-0.89	-0.21
SME_11	Hom-Tercil 1	0.27	Muj-Tercil 1	0.44	-0.17
SME_12	Hom-Tercil 1	-0.27	Muj-Tercil 1	-0.17	-0.09
SME_13	Hom-Tercil 1	0.31	Muj-Tercil 1	0.39	-0.09
SMF_09	Hom-Tercil 1	-0.61	Muj-Tercil 1	-0.46	-0.15
SMF_11	Hom-Tercil 1	0.11	Muj-Tercil 1	-0.2	0.31
SMF_12	Hom-Tercil 1	-0.1	Muj-Tercil 1	-0.14	0.04
SMF_13	Hom-Tercil 1	0.31	Muj-Tercil 1	0.8	-0.48
SMG_11	Hom-Tercil 1	-0.02	Muj-Tercil 1	0.04	-0.06
SMG_12	Hom-Tercil 1	-1.24	Muj-Tercil 1	-1.26	0.02
SMH_06	Hom-Tercil 1	-0.13	Muj-Tercil 1	-0.3	0.17
SMH_12	Hom-Tercil 1	-0.38	Muj-Tercil 1	-0.33	-0.04
SMH_13	Hom-Tercil 1	-0.7	Muj-Tercil 1	-0.98	0.28
SMA_09	Hom-Tercil 2	1.56	Muj-Tercil 2	1.41	0.14
SMA_10	Hom-Tercil 2	0.39	Muj-Tercil 2	0.26	0.13
SMA_11	Hom-Tercil 2	1.06	Muj-Tercil 2	1.08	-0.03
SMB_11	Hom-Tercil 2	-0.95	Muj-Tercil 2	-0.57	-0.38
SMB_12	Hom-Tercil 2	1.02	Muj-Tercil 2	1.32	-0.3
SMB_13	Hom-Tercil 2	1.32	Muj-Tercil 2	1.22	0.1
SMC_04	Hom-Tercil 2	-0.12	Muj-Tercil 2	-0.22	0.1
SMC_07	Hom-Tercil 2	0.47	Muj-Tercil 2	0.6	-0.14
SMC_08	Hom-Tercil 2	-2.75	Muj-Tercil 2	-3.04	0.29

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMC_09	Hom-Tercil 2	-0.17	Muj-Tercil 2	-0.19	0.03
SMC_10	Hom-Tercil 2	0.74	Muj-Tercil 2	1	-0.26
SMC_11	Hom-Tercil 2	-0.47	Muj-Tercil 2	-0.46	0
SMC_12	Hom-Tercil 2	-0.01	Muj-Tercil 2	-0.18	0.16
SMD_11	Hom-Tercil 2	0.26	Muj-Tercil 2	0.25	0.01
SME_09	Hom-Tercil 2	0.74	Muj-Tercil 2	0.93	-0.2
SME_10	Hom-Tercil 2	-1.25	Muj-Tercil 2	-0.92	-0.33
SME_11	Hom-Tercil 2	0.75	Muj-Tercil 2	0.67	0.08
SME_12	Hom-Tercil 2	-1.07	Muj-Tercil 2	-1.1	0.03
SME_13	Hom-Tercil 2	0.41	Muj-Tercil 2	0.11	0.3
SMF_09	Hom-Tercil 2	-1.26	Muj-Tercil 2	-0.8	-0.45
SMF_11	Hom-Tercil 2	0.1	Muj-Tercil 2	-0.1	0.19
SMF_12	Hom-Tercil 2	0.14	Muj-Tercil 2	0.16	-0.02
SMF_13	Hom-Tercil 2	1.2	Muj-Tercil 2	1.29	-0.08
SMG_11	Hom-Tercil 2	0.35	Muj-Tercil 2	0.17	0.18
SMG_12	Hom-Tercil 2	-1.64	Muj-Tercil 2	-1.46	-0.18
SMH_06	Hom-Tercil 2	0.39	Muj-Tercil 2	-0.05	0.45
SMH_12	Hom-Tercil 2	-0.56	Muj-Tercil 2	-0.4	-0.16
SMH_13	Hom-Tercil 2	-0.92	Muj-Tercil 2	-1.13	0.21
SMA_09	Hom-Tercil 3	1.67	Muj-Tercil 3	1.77	-0.1
SMA_10	Hom-Tercil 3	0.29	Muj-Tercil 3	0.28	0.01
SMA_11	Hom-Tercil 3	1.61	Muj-Tercil 3	1.32	0.28
SMB_11	Hom-Tercil 3	-1.34	Muj-Tercil 3	-0.86	-0.48
SMB_12	Hom-Tercil 3	1.64	Muj-Tercil 3	1.16	0.47
SMB_13	Hom-Tercil 3	1.43	Muj-Tercil 3	1.42	0.01
SMC_04	Hom-Tercil 3	-0.7	Muj-Tercil 3	-0.48	-0.22
SMC_07	Hom-Tercil 3	0.01	Muj-Tercil 3	0.62	-0.61
SMC_08	Hom-Tercil 3	-4.55>	Muj-Tercil 3	-4.55>	0
SMC_09	Hom-Tercil 3	-0.67	Muj-Tercil 3	-0.5	-0.17
SMC_10	Hom-Tercil 3	1.57	Muj-Tercil 3	1.93	-0.36
SMC_11	Hom-Tercil 3	-0.6	Muj-Tercil 3	-0.73	0.14
SMC_12	Hom-Tercil 3	-0.1	Muj-Tercil 3	-0.52	0.42
SMD_11	Hom-Tercil 3	0.52	Muj-Tercil 3	0.73	-0.21
SME_09	Hom-Tercil 3	1.46	Muj-Tercil 3	1.32	0.14
SME_10	Hom-Tercil 3	-2.11	Muj-Tercil 3	-1.64	-0.47
SME_11	Hom-Tercil 3	-0.09	Muj-Tercil 3	0.36	-0.45
SME_12	Hom-Tercil 3	-2.3	Muj-Tercil 3	-1.96	-0.34
SME_13	Hom-Tercil 3	0.59	Muj-Tercil 3	0.48	0.12
SMF_09	Hom-Tercil 3	-1.8	Muj-Tercil 3	-1.88	0.08
SMF_11	Hom-Tercil 3	-0.34	Muj-Tercil 3	-0.72	0.38
SMF_12	Hom-Tercil 3	-0.75	Muj-Tercil 3	-0.07	-0.69

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMF_13	Hom-Tercil 3	1.5	Muj-Tercil 3	1.68	-0.18
SMG_11	Hom-Tercil 3	-0.2	Muj-Tercil 3	-0.35	0.15
SMG_12	Hom-Tercil 3	-2.38	Muj-Tercil 3	-2.67	0.29
SMH_06	Hom-Tercil 3	0.25	Muj-Tercil 3	-0.35	0.59
SMH_12	Hom-Tercil 3	-0.72	Muj-Tercil 3	-0.56	-0.16
SMH_13	Hom-Tercil 3	-1.45	Muj-Tercil 3	-1.68	0.23

\* Valores positivos indican un DIF a favor de Mujeres; valores negativos indican un DIF a favor de Hombres. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 14. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_01	Bajo	0.64	Alto	0.71	-0.07
SMA_02	Bajo	-0.92	Alto	-1.19	0.26
SMA_05	Bajo	-0.19	Alto	0.05	-0.24
SMA_08	Bajo	-1.51	Alto	-1.85	0.35
SMA_12	Bajo	0.06	Alto	-0.39	0.45
SMB_01	Bajo	0.71	Alto	-0.27	0.98
SMB_02	Bajo	0.44	Alto	0.39	0.05
SMB_03	Bajo	-0.48	Alto	0.33	-0.81
SMB_04	Bajo	0.69	Alto	1.19	-0.49
SMB_05	Bajo	-1.26	Alto	-1.82	0.56
SMC_01	Bajo	-0.11	Alto	-0.08	-0.03
SMC_02	Bajo	-0.49	Alto	-0.34	-0.15
SMC_03	Bajo	0.4	Alto	0.51	-0.12
SMC_05	Bajo	1.43	Alto	1.66	-0.23
SMC_06	Bajo	0.17	Alto	-0.12	0.29
SMD_01	Bajo	0.69	Alto	1.14	-0.45
SMD_10	Bajo	-0.45	Alto	-0.77	0.32
SMD_12	Bajo	0.86	Alto	0.42	0.44
SME_01	Bajo	0.3	Alto	-0.07	0.36
SME_02	Bajo	-0.94	Alto	0.05	-0.99
SME_03	Bajo	-0.11	Alto	0.33	-0.44
SMF_01	Bajo	-0.48	Alto	-0.09	-0.39
SMF_02	Bajo	-1.16	Alto	-1.9	0.74
SMF_03	Bajo	0.51	Alto	0.54	-0.03
SMF_04	Bajo	-0.24	Alto	-0.15	-0.09
SMF_05	Bajo	0.03	Alto	-0.18	0.21
SMG_01	Bajo	0.88	Alto	1.29	-0.41
SMG_02	Bajo	-0.08	Alto	-0.08	0
SMG_03	Bajo	0.29	Alto	0.49	-0.2
SMG_04	Bajo	-0.92	Alto	-0.83	-0.09
SMG_05	Bajo	0.22	Alto	-0.07	0.29
SMG_08	Bajo	0.17	Alto	0.27	-0.1
SMG_09	Bajo	-0.75	Alto	-0.78	0.04
SMG_10	Bajo	-0.14	Alto	-0.28	0.14
SMH_01	Bajo	-0.79	Alto	-0.82	0.03
SMH_02	Bajo	0.42	Alto	0.37	0.05
SMH_03	Bajo	-0.16	Alto	0.24	-0.4

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMH_04	Bajo	1.24	Alto	1.64	-0.39
SMH_05	Bajo	-0.01	Alto	-0.29	0.28
SMH_07	Bajo	0.77	Alto	0.61	0.16

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 15. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_03	Bajo	0.15	Alto	0.02	0.13
SMA_04	Bajo	-0.09	Alto	-0.4	0.31
SMA_06	Bajo	-0.09	Alto	0.03	-0.12
SMA_07	Bajo	-0.04	Alto	-0.07	0.04
SMB_06	Bajo	-0.2	Alto	-0.07	-0.13
SMB_07	Bajo	-1.39	Alto	-1.28	-0.11
SMB_08	Bajo	0.48	Alto	0.89	-0.41
SMB_09	Bajo	-0.06	Alto	-0.72	0.66
SMB_10	Bajo	0.23	Alto	0.06	0.17
SMD_02	Bajo	1.54	Alto	1.22	0.32
SMD_03	Bajo	-0.68	Alto	-1.17	0.49
SMD_04	Bajo	0.28	Alto	1.08	-0.8
SMD_05	Bajo	-0.74	Alto	-0.08	-0.66
SMD_06	Bajo	1.26	Alto	0.83	0.43
SMD_07	Bajo	1.01	Alto	0.75	0.26
SMD_08	Bajo	-1.53	Alto	-1.51	-0.01
SMD_09	Bajo	-0.82	Alto	-1.25	0.43
SME_04	Bajo	0.67	Alto	0.96	-0.29
SME_05	Bajo	-0.48	Alto	-0.77	0.29
SME_06	Bajo	-0.86	Alto	-1.11	0.25
SME_07	Bajo	0.13	Alto	0.38	-0.25
SME_08	Bajo	-0.2	Alto	0.2	-0.4
SMF_06	Bajo	0.05	Alto	0.43	-0.38
SMF_07	Bajo	-0.25	Alto	-0.17	-0.08
SMF_08	Bajo	1.27	Alto	1.22	0.05
SMF_10	Bajo	-1.03	Alto	-1.23	0.2
SMG_06	Bajo	-1.07	Alto	-1.09	0.01
SMG_07	Bajo	2.96	Alto	2.15	0.81
SMH_08	Bajo	0.58	Alto	0.81	-0.22
SMH_09	Bajo	-0.66	Alto	-0.89	0.23
SMH_10	Bajo	0.09	Alto	-0.1	0.19
SMH_11	Bajo	-0.19	Alto	0.59	-0.78

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 16. Magnitud del Funcionamiento Diferencial de Ítems (DIF) Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_09	Bajo	1.38	Alto	1.56	-0.18
SMA_10	Bajo	0.32	Alto	0.26	0.06
SMA_11	Bajo	0.9	Alto	1.22	-0.32
SMB_11	Bajo	-0.29	Alto	-0.98	0.69
SMB_12	Bajo	0.79	Alto	1.28	-0.5
SMB_13	Bajo	1.07	Alto	1.39	-0.32
SMC_04	Bajo	0.09	Alto	-0.28	0.37
SMC_07	Bajo	0.45	Alto	0.4	0.05
SMC_08	Bajo	-1.98	Alto	-3.19	1.21
SMC_09	Bajo	0.03	Alto	-0.31	0.34
SMC_10	Bajo	0.24	Alto	1.06	-0.83
SMC_11	Bajo	-0.34	Alto	-0.44	0.1
SMC_12	Bajo	-0.32	Alto	0	-0.31
SMD_11	Bajo	-0.21	Alto	0.45	-0.65
SME_09	Bajo	0.47	Alto	1.22	-0.74
SME_10	Bajo	-1.07	Alto	-1.25	0.18
SME_11	Bajo	0.94	Alto	0.19	0.75
SME_12	Bajo	-0.44	Alto	-1.23	0.79
SME_13	Bajo	0.37	Alto	0.34	0.02
SMF_09	Bajo	-0.99	Alto	-0.95	-0.04
SMF_11	Bajo	-0.22	Alto	0.19	-0.4
SMF_12	Bajo	0.1	Alto	0.07	0.03
SMF_13	Bajo	0.99	Alto	1.33	-0.34
SMG_11	Bajo	0.1	Alto	0.22	-0.12
SMG_12	Bajo	-1.46	Alto	-1.57	0.11
SMH_06	Bajo	0.09	Alto	0.16	-0.07
SMH_12	Bajo	-0.49	Alto	-0.36	-0.13
SMH_13	Bajo	-0.87	Alto	-1.07	0.2

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 17. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Sentido Numérico y Pensamiento Algebraico del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_01	NSE Bajo-Tercil 1	0.44	NSE Alto-Tercil 1	0.77	-0.32
SMA_02	NSE Bajo-Tercil 1	-0.66	NSE Alto-Tercil 1	-0.59	-0.07
SMA_05	NSE Bajo-Tercil 1	-0.18	NSE Alto-Tercil 1	0.55	-0.73
SMA_08	NSE Bajo-Tercil 1	-1.44	NSE Alto-Tercil 1	-1.69	0.25
SMA_12	NSE Bajo-Tercil 1	0.19	NSE Alto-Tercil 1	0.03	0.16
SMB_01	NSE Bajo-Tercil 1	0.9	NSE Alto-Tercil 1	0.51	0.39
SMB_02	NSE Bajo-Tercil 1	0.39	NSE Alto-Tercil 1	0.2	0.19
SMB_03	NSE Bajo-Tercil 1	-0.71	NSE Alto-Tercil 1	-0.68	-0.03
SMB_04	NSE Bajo-Tercil 1	0.61	NSE Alto-Tercil 1	0.32	0.28
SMB_05	NSE Bajo-Tercil 1	-1.13	NSE Alto-Tercil 1	-1.39	0.26
SMC_01	NSE Bajo-Tercil 1	-0.34	NSE Alto-Tercil 1	0.09	-0.42
SMC_02	NSE Bajo-Tercil 1	-0.45	NSE Alto-Tercil 1	-0.5	0.05
SMC_03	NSE Bajo-Tercil 1	0.43	NSE Alto-Tercil 1	0.45	-0.03
SMC_05	NSE Bajo-Tercil 1	1.17	NSE Alto-Tercil 1	0.88	0.29
SMC_06	NSE Bajo-Tercil 1	0.54	NSE Alto-Tercil 1	0.35	0.19
SMD_01	NSE Bajo-Tercil 1	0.55	NSE Alto-Tercil 1	0.43	0.12
SMD_10	NSE Bajo-Tercil 1	-0.19	NSE Alto-Tercil 1	-0.44	0.25
SMD_12	NSE Bajo-Tercil 1	0.64	NSE Alto-Tercil 1	0.61	0.03
SME_01	NSE Bajo-Tercil 1	0.44	NSE Alto-Tercil 1	0.57	-0.14
SME_02	NSE Bajo-Tercil 1	-1.14	NSE Alto-Tercil 1	-0.58	-0.56
SME_03	NSE Bajo-Tercil 1	-0.13	NSE Alto-Tercil 1	-0.03	-0.1
SMF_01	NSE Bajo-Tercil 1	-0.74	NSE Alto-Tercil 1	-0.5	-0.24
SMF_02	NSE Bajo-Tercil 1	-0.84	NSE Alto-Tercil 1	-1.51	0.67
SMF_03	NSE Bajo-Tercil 1	0.59	NSE Alto-Tercil 1	0.44	0.14
SMF_04	NSE Bajo-Tercil 1	-0.4	NSE Alto-Tercil 1	-0.51	0.11
SMF_05	NSE Bajo-Tercil 1	0.08	NSE Alto-Tercil 1	-0.26	0.34
SMG_01	NSE Bajo-Tercil 1	0.29	NSE Alto-Tercil 1	0.87	-0.57
SMG_02	NSE Bajo-Tercil 1	-0.36	NSE Alto-Tercil 1	0.77	-1.12
SMG_03	NSE Bajo-Tercil 1	0.07	NSE Alto-Tercil 1	0.18	-0.11
SMG_04	NSE Bajo-Tercil 1	-0.87	NSE Alto-Tercil 1	-0.53	-0.34
SMG_05	NSE Bajo-Tercil 1	0.62	NSE Alto-Tercil 1	-0.07	0.69
SMG_08	NSE Bajo-Tercil 1	0.37	NSE Alto-Tercil 1	0.64	-0.26
SMG_09	NSE Bajo-Tercil 1	-0.75	NSE Alto-Tercil 1	-0.77	0.03
SMG_10	NSE Bajo-Tercil 1	0.12	NSE Alto-Tercil 1	-0.18	0.29
SMH_01	NSE Bajo-Tercil 1	-0.98	NSE Alto-Tercil 1	-1.16	0.18
SMH_02	NSE Bajo-Tercil 1	0.48	NSE Alto-Tercil 1	0.11	0.37
SMH_03	NSE Bajo-Tercil 1	-0.28	NSE Alto-Tercil 1	-0.01	-0.27
SMH_04	NSE Bajo-Tercil 1	0.52	NSE Alto-Tercil 1	0.83	-0.31
SMH_05	NSE Bajo-Tercil 1	0.07	NSE Alto-Tercil 1	-0.2	0.27
SMH_07	NSE Bajo-Tercil 1	0.4	NSE Alto-Tercil 1	0.57	-0.17



Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_01	NSE Bajo-Tercil 2	0.68	NSE Alto-Tercil 2	0.77	-0.09
SMA_02	NSE Bajo-Tercil 2	-1.06	NSE Alto-Tercil 2	-1.31	0.24
SMA_05	NSE Bajo-Tercil 2	-0.2	NSE Alto-Tercil 2	0.08	-0.27
SMA_08	NSE Bajo-Tercil 2	-1.55	NSE Alto-Tercil 2	-1.93	0.38
SMA_12	NSE Bajo-Tercil 2	0.02	NSE Alto-Tercil 2	-0.43	0.44
SMB_01	NSE Bajo-Tercil 2	0.69	NSE Alto-Tercil 2	-0.28	0.97
SMB_02	NSE Bajo-Tercil 2	0.47	NSE Alto-Tercil 2	0.39	0.08
SMB_03	NSE Bajo-Tercil 2	-0.4	NSE Alto-Tercil 2	0.24	-0.63
SMB_04	NSE Bajo-Tercil 2	0.68	NSE Alto-Tercil 2	1.11	-0.42
SMB_05	NSE Bajo-Tercil 2	-1.32	NSE Alto-Tercil 2	-1.91	0.59
SMC_01	NSE Bajo-Tercil 2	-0.02	NSE Alto-Tercil 2	-0.08	0.06
SMC_02	NSE Bajo-Tercil 2	-0.5	NSE Alto-Tercil 2	-0.4	-0.1
SMC_03	NSE Bajo-Tercil 2	0.37	NSE Alto-Tercil 2	0.51	-0.14
SMC_05	NSE Bajo-Tercil 2	1.47	NSE Alto-Tercil 2	1.76	-0.29
SMC_06	NSE Bajo-Tercil 2	0.08	NSE Alto-Tercil 2	-0.14	0.23
SMD_01	NSE Bajo-Tercil 2	0.67	NSE Alto-Tercil 2	1.16	-0.49
SMD_10	NSE Bajo-Tercil 2	-0.53	NSE Alto-Tercil 2	-0.84	0.31
SMD_12	NSE Bajo-Tercil 2	0.96	NSE Alto-Tercil 2	0.55	0.41
SME_01	NSE Bajo-Tercil 2	0.27	NSE Alto-Tercil 2	-0.1	0.37
SME_02	NSE Bajo-Tercil 2	-0.87	NSE Alto-Tercil 2	0.1	-0.97
SME_03	NSE Bajo-Tercil 2	-0.13	NSE Alto-Tercil 2	0.28	-0.41
SMF_01	NSE Bajo-Tercil 2	-0.38	NSE Alto-Tercil 2	-0.07	-0.31
SMF_02	NSE Bajo-Tercil 2	-1.34	NSE Alto-Tercil 2	-2.01	0.68
SMF_03	NSE Bajo-Tercil 2	0.5	NSE Alto-Tercil 2	0.51	-0.01
SMF_04	NSE Bajo-Tercil 2	-0.18	NSE Alto-Tercil 2	-0.07	-0.11
SMF_05	NSE Bajo-Tercil 2	0.02	NSE Alto-Tercil 2	-0.19	0.21
SMG_01	NSE Bajo-Tercil 2	0.96	NSE Alto-Tercil 2	1.15	-0.19
SMG_02	NSE Bajo-Tercil 2	0.02	NSE Alto-Tercil 2	-0.12	0.14
SMG_03	NSE Bajo-Tercil 2	0.34	NSE Alto-Tercil 2	0.49	-0.14
SMG_04	NSE Bajo-Tercil 2	-0.93	NSE Alto-Tercil 2	-0.83	-0.1
SMG_05	NSE Bajo-Tercil 2	0.17	NSE Alto-Tercil 2	-0.03	0.2
SMG_08	NSE Bajo-Tercil 2	0.15	NSE Alto-Tercil 2	0.3	-0.16
SMG_09	NSE Bajo-Tercil 2	-0.75	NSE Alto-Tercil 2	-0.75	0
SMG_10	NSE Bajo-Tercil 2	-0.18	NSE Alto-Tercil 2	-0.22	0.05
SMH_01	NSE Bajo-Tercil 2	-0.72	NSE Alto-Tercil 2	-0.76	0.04
SMH_02	NSE Bajo-Tercil 2	0.37	NSE Alto-Tercil 2	0.35	0.02
SMH_03	NSE Bajo-Tercil 2	-0.12	NSE Alto-Tercil 2	0.28	-0.39
SMH_04	NSE Bajo-Tercil 2	1.41	NSE Alto-Tercil 2	1.66	-0.24
SMH_05	NSE Bajo-Tercil 2	-0.03	NSE Alto-Tercil 2	-0.26	0.24
SMH_07	NSE Bajo-Tercil 2	0.86	NSE Alto-Tercil 2	0.7	0.17
SMA_01	NSE Bajo-Tercil 3	0.98	NSE Alto-Tercil 3	0.32	0.66
SMA_02	NSE Bajo-Tercil 3	-5.37	NSE Alto-Tercil 3	-2.08	-3.29
SMA_05	NSE Bajo-Tercil 3	-0.02	NSE Alto-Tercil 3	-0.83	0.81
SMA_08	NSE Bajo-Tercil 3	-5.37	NSE Alto-Tercil 3	-1.64	-3.74

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_12	NSE Bajo-Tercil 3	0.11	NSE Alto-Tercil 3	-0.69	0.8
SMB_01	NSE Bajo-Tercil 3	-0.04	NSE Alto-Tercil 3	-1.47	1.43
SMB_02	NSE Bajo-Tercil 3	-0.95	NSE Alto-Tercil 3	0.39	-1.34
SMB_03	NSE Bajo-Tercil 3	-0.13	NSE Alto-Tercil 3	1.75	-1.88
SMB_04	NSE Bajo-Tercil 3	1.9	NSE Alto-Tercil 3	1.87	0.04
SMB_05	NSE Bajo-Tercil 3	-4.71	NSE Alto-Tercil 3	-7.39	2.68
SMC_01	NSE Bajo-Tercil 3	0.21	NSE Alto-Tercil 3	-0.44	0.66
SMC_02	NSE Bajo-Tercil 3	-0.4	NSE Alto-Tercil 3	0.73	-1.13
SMC_03	NSE Bajo-Tercil 3	1.45	NSE Alto-Tercil 3	0.65	0.8
SMC_05	NSE Bajo-Tercil 3	1.86	NSE Alto-Tercil 3	1.51	0.35
SMC_06	NSE Bajo-Tercil 3	-1.08	NSE Alto-Tercil 3	-0.94	-0.15
SMD_01	NSE Bajo-Tercil 3	1.76	NSE Alto-Tercil 3	1.31	0.46
SMD_10	NSE Bajo-Tercil 3	-1.21	NSE Alto-Tercil 3	-0.57	-0.64
SMD_12	NSE Bajo-Tercil 3	-0.49	NSE Alto-Tercil 3	-0.91	0.42
SME_01	NSE Bajo-Tercil 3	0.12	NSE Alto-Tercil 3	-0.51	0.63
SME_02	NSE Bajo-Tercil 3	0.04	NSE Alto-Tercil 3	0.32	-0.28
SME_03	NSE Bajo-Tercil 3	1.22	NSE Alto-Tercil 3	1.15	0.07
SMF_01	NSE Bajo-Tercil 3	0.64	NSE Alto-Tercil 3	0.51	0.13
SMF_02	NSE Bajo-Tercil 3	-0.93	NSE Alto-Tercil 3	-3.88	2.95
SMF_03	NSE Bajo-Tercil 3	-0.16	NSE Alto-Tercil 3	0.95	-1.11
SMF_04	NSE Bajo-Tercil 3	-0.15	NSE Alto-Tercil 3	-1.06	0.91
SMF_05	NSE Bajo-Tercil 3	0.04	NSE Alto-Tercil 3	0.18	-0.14
SMG_01	NSE Bajo-Tercil 3	2.24	NSE Alto-Tercil 3	2.12	0.11
SMG_02	NSE Bajo-Tercil 3	-0.41	NSE Alto-Tercil 3	-0.56	0.14
SMG_03	NSE Bajo-Tercil 3	0.17	NSE Alto-Tercil 3	0.71	-0.54
SMG_04	NSE Bajo-Tercil 3	-3.33	NSE Alto-Tercil 3	-1.8	-1.53
SMG_05	NSE Bajo-Tercil 3	-1.23	NSE Alto-Tercil 3	-0.52	-0.72
SMG_08	NSE Bajo-Tercil 3	-0.19	NSE Alto-Tercil 3	-0.46	0.27
SMG_09	NSE Bajo-Tercil 3	-2.77	NSE Alto-Tercil 3	-1.97	-0.8
SMG_10	NSE Bajo-Tercil 3	-3.03	NSE Alto-Tercil 3	-1.76	-1.26
SMH_01	NSE Bajo-Tercil 3	-1.88	NSE Alto-Tercil 3	-1.25	-0.63
SMH_02	NSE Bajo-Tercil 3	1.29	NSE Alto-Tercil 3	0.84	0.45
SMH_03	NSE Bajo-Tercil 3	-1.05	NSE Alto-Tercil 3	-0.02	-1.04
SMH_04	NSE Bajo-Tercil 3	1.68	NSE Alto-Tercil 3	1.78	-0.1
SMH_05	NSE Bajo-Tercil 3	-0.29	NSE Alto-Tercil 3	-1.18	0.9
SMH_07	NSE Bajo-Tercil 3	0.41	NSE Alto-Tercil 3	-0.4	0.81

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 18. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Forma, Espacio y Medida del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_03	NSE Bajo-Tercil 1	0.03	NSE Alto-Tercil 1	0.14	-0.11
SMA_04	NSE Bajo-Tercil 1	0.28	NSE Alto-Tercil 1	0.05	0.23
SMA_06	NSE Bajo-Tercil 1	-0.27	NSE Alto-Tercil 1	-0.22	-0.05
SMA_07	NSE Bajo-Tercil 1	0.23	NSE Alto-Tercil 1	-0.33	0.56
SMB_06	NSE Bajo-Tercil 1	-0.44	NSE Alto-Tercil 1	0.06	-0.5
SMB_07	NSE Bajo-Tercil 1	-1.21	NSE Alto-Tercil 1	-1.15	-0.07
SMB_08	NSE Bajo-Tercil 1	0.51	NSE Alto-Tercil 1	0.54	-0.03
SMB_09	NSE Bajo-Tercil 1	-0.05	NSE Alto-Tercil 1	-0.61	0.55
SMB_10	NSE Bajo-Tercil 1	0.03	NSE Alto-Tercil 1	0.04	-0.01
SMD_02	NSE Bajo-Tercil 1	1.36	NSE Alto-Tercil 1	1.23	0.13
SMD_03	NSE Bajo-Tercil 1	-0.43	NSE Alto-Tercil 1	-1.04	0.6
SMD_04	NSE Bajo-Tercil 1	-0.05	NSE Alto-Tercil 1	-0.15	0.1
SMD_05	NSE Bajo-Tercil 1	-0.92	NSE Alto-Tercil 1	-0.65	-0.27
SMD_06	NSE Bajo-Tercil 1	0.94	NSE Alto-Tercil 1	1.69	-0.75
SMD_07	NSE Bajo-Tercil 1	0.97	NSE Alto-Tercil 1	0.35	0.63
SMD_08	NSE Bajo-Tercil 1	-1.33	NSE Alto-Tercil 1	-1.19	-0.14
SMD_09	NSE Bajo-Tercil 1	-0.51	NSE Alto-Tercil 1	-0.86	0.35
SME_04	NSE Bajo-Tercil 1	0.19	NSE Alto-Tercil 1	0.85	-0.66
SME_05	NSE Bajo-Tercil 1	-0.51	NSE Alto-Tercil 1	-1.05	0.54
SME_06	NSE Bajo-Tercil 1	-0.6	NSE Alto-Tercil 1	-0.75	0.14
SME_07	NSE Bajo-Tercil 1	0.23	NSE Alto-Tercil 1	0.45	-0.22
SME_08	NSE Bajo-Tercil 1	-0.43	NSE Alto-Tercil 1	-0.03	-0.4
SMF_06	NSE Bajo-Tercil 1	-0.2	NSE Alto-Tercil 1	0.42	-0.61
SMF_07	NSE Bajo-Tercil 1	-0.34	NSE Alto-Tercil 1	-0.32	-0.02
SMF_08	NSE Bajo-Tercil 1	0.88	NSE Alto-Tercil 1	1.86	-0.98
SMF_10	NSE Bajo-Tercil 1	-0.9	NSE Alto-Tercil 1	-1.03	0.13
SMG_06	NSE Bajo-Tercil 1	-0.93	NSE Alto-Tercil 1	-0.57	-0.35
SMG_07	NSE Bajo-Tercil 1	2.99	NSE Alto-Tercil 1	2.94	0.05
SMH_08	NSE Bajo-Tercil 1	0.59	NSE Alto-Tercil 1	0.33	0.25
SMH_09	NSE Bajo-Tercil 1	-0.68	NSE Alto-Tercil 1	-1.05	0.37
SMH_10	NSE Bajo-Tercil 1	0.13	NSE Alto-Tercil 1	0.67	-0.54
SMH_11	NSE Bajo-Tercil 1	-0.47	NSE Alto-Tercil 1	0.17	-0.65
SMA_03	NSE Bajo-Tercil 2	0.2	NSE Alto-Tercil 2	0.02	0.18
SMA_04	NSE Bajo-Tercil 2	-0.22	NSE Alto-Tercil 2	-0.47	0.26
SMA_06	NSE Bajo-Tercil 2	-0.04	NSE Alto-Tercil 2	0.04	-0.09
SMA_07	NSE Bajo-Tercil 2	-0.13	NSE Alto-Tercil 2	-0.02	-0.11
SMB_06	NSE Bajo-Tercil 2	-0.05	NSE Alto-Tercil 2	-0.07	0.02

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMB_07	NSE Bajo-Tercil 2	-1.6	NSE Alto-Tercil 2	-1.31	-0.29
SMB_08	NSE Bajo-Tercil 2	0.47	NSE Alto-Tercil 2	0.87	-0.4
SMB_09	NSE Bajo-Tercil 2	-0.05	NSE Alto-Tercil 2	-0.75	0.69
SMB_10	NSE Bajo-Tercil 2	0.33	NSE Alto-Tercil 2	0.1	0.23
SMD_02	NSE Bajo-Tercil 2	1.62	NSE Alto-Tercil 2	1.28	0.33
SMD_03	NSE Bajo-Tercil 2	-0.82	NSE Alto-Tercil 2	-1.2	0.38
SMD_04	NSE Bajo-Tercil 2	0.35	NSE Alto-Tercil 2	1.05	-0.71
SMD_05	NSE Bajo-Tercil 2	-0.64	NSE Alto-Tercil 2	0.03	-0.67
SMD_06	NSE Bajo-Tercil 2	1.43	NSE Alto-Tercil 2	0.83	0.61
SMD_07	NSE Bajo-Tercil 2	1.04	NSE Alto-Tercil 2	0.84	0.2
SMD_08	NSE Bajo-Tercil 2	-1.69	NSE Alto-Tercil 2	-1.6	-0.1
SMD_09	NSE Bajo-Tercil 2	-0.98	NSE Alto-Tercil 2	-1.34	0.36
SME_04	NSE Bajo-Tercil 2	0.82	NSE Alto-Tercil 2	0.93	-0.12
SME_05	NSE Bajo-Tercil 2	-0.47	NSE Alto-Tercil 2	-0.7	0.23
SME_06	NSE Bajo-Tercil 2	-0.99	NSE Alto-Tercil 2	-1.2	0.2
SME_07	NSE Bajo-Tercil 2	0.09	NSE Alto-Tercil 2	0.38	-0.28
SME_08	NSE Bajo-Tercil 2	-0.11	NSE Alto-Tercil 2	0.21	-0.32
SMF_06	NSE Bajo-Tercil 2	0.12	NSE Alto-Tercil 2	0.39	-0.27
SMF_07	NSE Bajo-Tercil 2	-0.21	NSE Alto-Tercil 2	-0.19	-0.03
SMF_08	NSE Bajo-Tercil 2	1.33	NSE Alto-Tercil 2	1.22	0.12
SMF_10	NSE Bajo-Tercil 2	-1.08	NSE Alto-Tercil 2	-1.26	0.18
SMG_06	NSE Bajo-Tercil 2	-1.14	NSE Alto-Tercil 2	-1.23	0.09
SMG_07	NSE Bajo-Tercil 2	3.11	NSE Alto-Tercil 2	2.32	0.79
SMH_08	NSE Bajo-Tercil 2	0.55	NSE Alto-Tercil 2	0.86	-0.31
SMH_09	NSE Bajo-Tercil 2	-0.64	NSE Alto-Tercil 2	-0.81	0.16
SMH_10	NSE Bajo-Tercil 2	0.08	NSE Alto-Tercil 2	-0.2	0.27
SMH_11	NSE Bajo-Tercil 2	-0.09	NSE Alto-Tercil 2	0.57	-0.65
SMA_03	NSE Bajo-Tercil 3	-0.82	NSE Alto-Tercil 3	-0.83	0.01
SMA_04	NSE Bajo-Tercil 3	-0.09	NSE Alto-Tercil 3	-1.91	1.83
SMA_06	NSE Bajo-Tercil 3	1.2	NSE Alto-Tercil 3	0.68	0.52
SMA_07	NSE Bajo-Tercil 3	-1.58	NSE Alto-Tercil 3	0.08	-1.66
SMB_06	NSE Bajo-Tercil 3	-0.55	NSE Alto-Tercil 3	-0.6	0.04
SMB_07	NSE Bajo-Tercil 3	-0.72	NSE Alto-Tercil 3	-1.45	0.73
SMB_08	NSE Bajo-Tercil 3	0.57	NSE Alto-Tercil 3	1.47	-0.89
SMB_09	NSE Bajo-Tercil 3	-0.96	NSE Alto-Tercil 3	-1.03	0.06
SMB_10	NSE Bajo-Tercil 3	0.48	NSE Alto-Tercil 3	-0.47	0.95
SMD_02	NSE Bajo-Tercil 3	0.82	NSE Alto-Tercil 3	0.75	0.07
SMD_03	NSE Bajo-Tercil 3	0.03	NSE Alto-Tercil 3	-1.34	1.37
SMD_04	NSE Bajo-Tercil 3	2.38	NSE Alto-Tercil 3	2.31	0.06
SMD_05	NSE Bajo-Tercil 3	-1.76	NSE Alto-Tercil 3	-0.56	-1.2
SMD_06	NSE Bajo-Tercil 3	-1.48	NSE Alto-Tercil 3	0.39	-1.88

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMD_07	NSE Bajo-Tercil 3	0.5	NSE Alto-Tercil 3	0.32	0.18
SMD_08	NSE Bajo-Tercil 3	-1.42	NSE Alto-Tercil 3	-2.78	1.36
SMD_09	NSE Bajo-Tercil 3	-4.78>	NSE Alto-Tercil 3	-1.9	-2.89
SME_04	NSE Bajo-Tercil 3	1.54	NSE Alto-Tercil 3	1.35	0.2
SME_05	NSE Bajo-Tercil 3	-0.23	NSE Alto-Tercil 3	-1.15	0.92
SME_06	NSE Bajo-Tercil 3	-2.5	NSE Alto-Tercil 3	-1.09	-1.41
SME_07	NSE Bajo-Tercil 3	0.48	NSE Alto-Tercil 3	0.39	0.09
SME_08	NSE Bajo-Tercil 3	0.76	NSE Alto-Tercil 3	0.61	0.14
SMF_06	NSE Bajo-Tercil 3	0.35	NSE Alto-Tercil 3	1.19	-0.84
SMF_07	NSE Bajo-Tercil 3	-0.88	NSE Alto-Tercil 3	0.76	-1.64
SMF_08	NSE Bajo-Tercil 3	1.08	NSE Alto-Tercil 3	0.91	0.16
SMF_10	NSE Bajo-Tercil 3	-2.81	NSE Alto-Tercil 3	-1.87	-0.95
SMG_06	NSE Bajo-Tercil 3	-5.42>	NSE Alto-Tercil 3	-1.22	-4.2
SMG_07	NSE Bajo-Tercil 3	2.01	NSE Alto-Tercil 3	1.29	0.72
SMH_08	NSE Bajo-Tercil 3	1.93	NSE Alto-Tercil 3	1.08	0.86
SMH_09	NSE Bajo-Tercil 3	-3.61	NSE Alto-Tercil 3	-2.4	-1.21
SMH_10	NSE Bajo-Tercil 3	-0.19	NSE Alto-Tercil 3	-1.31	1.12
SMH_11	NSE Bajo-Tercil 3	0.71	NSE Alto-Tercil 3	1.78	-1.07

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.

**Anexo 19. Magnitud del Funcionamiento Diferencial de Ítems (DIF) No Uniforme en la subescala Manejo de la Información del Excale 09 de Matemáticas en 2012, desglosado por Nivel Socioeconómico**

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMA_09	NSE Bajo-Tercil 1	1.4	NSE Alto-Tercil 1	1.57	-0.17
SMA_10	NSE Bajo-Tercil 1	0	NSE Alto-Tercil 1	-0.41	0.41
SMA_11	NSE Bajo-Tercil 1	0.42	NSE Alto-Tercil 1	0.75	-0.33
SMB_11	NSE Bajo-Tercil 1	-0.07	NSE Alto-Tercil 1	-0.56	0.49
SMB_12	NSE Bajo-Tercil 1	0.42	NSE Alto-Tercil 1	1.08	-0.66
SMB_13	NSE Bajo-Tercil 1	0.88	NSE Alto-Tercil 1	1.1	-0.22
SMC_04	NSE Bajo-Tercil 1	-0.06	NSE Alto-Tercil 1	0.11	-0.17
SMC_07	NSE Bajo-Tercil 1	0.53	NSE Alto-Tercil 1	0.7	-0.17
SMC_08	NSE Bajo-Tercil 1	-1.76	NSE Alto-Tercil 1	-2.71	0.95
SMC_09	NSE Bajo-Tercil 1	0.42	NSE Alto-Tercil 1	0.13	0.29
SMC_10	NSE Bajo-Tercil 1	-0.52	NSE Alto-Tercil 1	-0.13	-0.39
SMC_11	NSE Bajo-Tercil 1	-0.31	NSE Alto-Tercil 1	0.05	-0.36
SMC_12	NSE Bajo-Tercil 1	-0.58	NSE Alto-Tercil 1	-0.08	-0.5
SMD_11	NSE Bajo-Tercil 1	-0.54	NSE Alto-Tercil 1	-0.24	-0.3
SME_09	NSE Bajo-Tercil 1	0.17	NSE Alto-Tercil 1	0.76	-0.6
SME_10	NSE Bajo-Tercil 1	-1.05	NSE Alto-Tercil 1	-1.11	0.06
SME_11	NSE Bajo-Tercil 1	0.47	NSE Alto-Tercil 1	-0.27	0.73
SME_12	NSE Bajo-Tercil 1	0.35	NSE Alto-Tercil 1	0.06	0.3
SME_13	NSE Bajo-Tercil 1	0.43	NSE Alto-Tercil 1	0.61	-0.17
SMF_09	NSE Bajo-Tercil 1	-0.56	NSE Alto-Tercil 1	-0.41	-0.15
SMF_11	NSE Bajo-Tercil 1	-0.18	NSE Alto-Tercil 1	0.31	-0.49
SMF_12	NSE Bajo-Tercil 1	-0.1	NSE Alto-Tercil 1	-0.17	0.08
SMF_13	NSE Bajo-Tercil 1	0.49	NSE Alto-Tercil 1	0.97	-0.49
SMG_11	NSE Bajo-Tercil 1	-0.07	NSE Alto-Tercil 1	0.23	-0.29
SMG_12	NSE Bajo-Tercil 1	-1.34	NSE Alto-Tercil 1	-1.4	0.06
SMH_06	NSE Bajo-Tercil 1	-0.25	NSE Alto-Tercil 1	-0.14	-0.1
SMH_12	NSE Bajo-Tercil 1	-0.47	NSE Alto-Tercil 1	-0.21	-0.26
SMH_13	NSE Bajo-Tercil 1	-0.73	NSE Alto-Tercil 1	-1.22	0.48
SMA_09	NSE Bajo-Tercil 2	1.39	NSE Alto-Tercil 2	1.45	-0.06
SMA_10	NSE Bajo-Tercil 2	0.43	NSE Alto-Tercil 2	0.39	0.04
SMA_11	NSE Bajo-Tercil 2	1.04	NSE Alto-Tercil 2	1.12	-0.08
SMB_11	NSE Bajo-Tercil 2	-0.37	NSE Alto-Tercil 2	-1.03	0.67
SMB_12	NSE Bajo-Tercil 2	0.88	NSE Alto-Tercil 2	1.18	-0.3
SMB_13	NSE Bajo-Tercil 2	1.08	NSE Alto-Tercil 2	1.45	-0.37
SMC_04	NSE Bajo-Tercil 2	0.16	NSE Alto-Tercil 2	-0.31	0.47
SMC_07	NSE Bajo-Tercil 2	0.43	NSE Alto-Tercil 2	0.43	-0.01
SMC_08	NSE Bajo-Tercil 2	-2.2	NSE Alto-Tercil 2	-3.51	1.31

Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMC_09	NSE Bajo-Tercil 2	-0.07	NSE Alto-Tercil 2	-0.37	0.3
SMC_10	NSE Bajo-Tercil 2	0.5	NSE Alto-Tercil 2	1.04	-0.54
SMC_11	NSE Bajo-Tercil 2	-0.35	NSE Alto-Tercil 2	-0.47	0.12
SMC_12	NSE Bajo-Tercil 2	-0.21	NSE Alto-Tercil 2	0.05	-0.26
SMD_11	NSE Bajo-Tercil 2	-0.11	NSE Alto-Tercil 2	0.45	-0.56
SME_09	NSE Bajo-Tercil 2	0.53	NSE Alto-Tercil 2	1.11	-0.58
SME_10	NSE Bajo-Tercil 2	-1.07	NSE Alto-Tercil 2	-1.23	0.15
SME_11	NSE Bajo-Tercil 2	1.16	NSE Alto-Tercil 2	0.34	0.83
SME_12	NSE Bajo-Tercil 2	-0.75	NSE Alto-Tercil 2	-1.43	0.68
SME_13	NSE Bajo-Tercil 2	0.35	NSE Alto-Tercil 2	0.31	0.04
SMF_09	NSE Bajo-Tercil 2	-1.23	NSE Alto-Tercil 2	-1	-0.23
SMF_11	NSE Bajo-Tercil 2	-0.23	NSE Alto-Tercil 2	0.27	-0.5
SMF_12	NSE Bajo-Tercil 2	0.18	NSE Alto-Tercil 2	0.18	0
SMF_13	NSE Bajo-Tercil 2	1.06	NSE Alto-Tercil 2	1.28	-0.23
SMG_11	NSE Bajo-Tercil 2	0.2	NSE Alto-Tercil 2	0.31	-0.11
SMG_12	NSE Bajo-Tercil 2	-1.56	NSE Alto-Tercil 2	-1.57	0.01
SMH_06	NSE Bajo-Tercil 2	0.19	NSE Alto-Tercil 2	0.17	0.02
SMH_12	NSE Bajo-Tercil 2	-0.49	NSE Alto-Tercil 2	-0.31	-0.18
SMH_13	NSE Bajo-Tercil 2	-0.96	NSE Alto-Tercil 2	-1	0.04
SMA_09	NSE Bajo-Tercil 3	1.29	NSE Alto-Tercil 3	1.81	-0.51
SMA_10	NSE Bajo-Tercil 3	0.26	NSE Alto-Tercil 3	0.13	0.13
SMA_11	NSE Bajo-Tercil 3	1.03	NSE Alto-Tercil 3	1.63	-0.6
SMB_11	NSE Bajo-Tercil 3	-1.03	NSE Alto-Tercil 3	-1.94	0.91
SMB_12	NSE Bajo-Tercil 3	1.17	NSE Alto-Tercil 3	1.62	-0.46
SMB_13	NSE Bajo-Tercil 3	1.4	NSE Alto-Tercil 3	1.34	0.06
SMC_04	NSE Bajo-Tercil 3	-0.46	NSE Alto-Tercil 3	-0.45	-0.01
SMC_07	NSE Bajo-Tercil 3	0.45	NSE Alto-Tercil 3	-0.03	0.48
SMC_08	NSE Bajo-Tercil 3	-5.43>	NSE Alto-Tercil 3	-7.91>	2.48
SMC_09	NSE Bajo-Tercil 3	-0.72	NSE Alto-Tercil 3	-0.19	-0.53
SMC_10	NSE Bajo-Tercil 3	1.58	NSE Alto-Tercil 3	1.73	-0.15
SMC_11	NSE Bajo-Tercil 3	-0.65	NSE Alto-Tercil 3	-0.97	0.31
SMC_12	NSE Bajo-Tercil 3	-0.37	NSE Alto-Tercil 3	-0.47	0.1
SMD_11	NSE Bajo-Tercil 3	0.79	NSE Alto-Tercil 3	0.9	-0.11
SME_09	NSE Bajo-Tercil 3	1.57	NSE Alto-Tercil 3	1.66	-0.08
SME_10	NSE Bajo-Tercil 3	-0.89	NSE Alto-Tercil 3	-2.51	1.62
SME_11	NSE Bajo-Tercil 3	0.4	NSE Alto-Tercil 3	-0.34	0.75
SME_12	NSE Bajo-Tercil 3	-6.14>	NSE Alto-Tercil 3	-3.08	-3.06
SME_13	NSE Bajo-Tercil 3	0.18	NSE Alto-Tercil 3	0.42	-0.24
SMF_09	NSE Bajo-Tercil 3	-1.48	NSE Alto-Tercil 3	-2.1	0.62
SMF_11	NSE Bajo-Tercil 3	-0.4	NSE Alto-Tercil 3	-0.48	0.08
SMF_12	NSE Bajo-Tercil 3	-0.21	NSE Alto-Tercil 3	-0.53	0.31



Reactivo	Alumno-Categoría	Nivel de dificultad	Alumno-Categoría	Nivel de dificultad	Magnitud de DIF*
SMF_13	NSE Bajo-Tercil 3	2.02	NSE Alto-Tercil 3	1.56	0.46
SMG_11	NSE Bajo-Tercil 3	-0.43	NSE Alto-Tercil 3	-0.18	-0.24
SMG_12	NSE Bajo-Tercil 3	-2.48	NSE Alto-Tercil 3	-2.52	0.04
SMH_06	NSE Bajo-Tercil 3	-0.01	NSE Alto-Tercil 3	0.27	-0.29
SMH_12	NSE Bajo-Tercil 3	-0.72	NSE Alto-Tercil 3	-0.88	0.16
SMH_13	NSE Bajo-Tercil 3	-0.3	NSE Alto-Tercil 3	-1.52	1.23

\* Valores positivos indican un DIF a favor de alumnos con un NSE Alto; valores negativos indican un DIF a favor de alumnos con un NSE Bajo. Celdas en amarillo advierten un DIF moderado ( $\geq 0.43$  &  $< 0.64$ ) y celdas en rojo un DIF enorme ( $\geq 0.64$ ).

**Fuente:** elaboración propia.



**Anexo 20. Resumen de reactivos con DIF (uniforme y/o no uniforme) del Excale 09 de Matemáticas en 2012**

Dominio	Reactivo	SEXO					NIVEL SOCIOECONÓMICO				
		DIF Uniforme	DIF No Uniforme			SELEC	DIF Uniforme	DIF No Uniforme			SELEC
			1	2	3			1	2	3	
SNPA	SMA_01				-					++	
SNPA	SMA_02		-							--	
SNPA	SMA_05							--		++	1
SNPA	SMA_08				--					--	
SNPA	SMA_12						+		+	++	
SNPA	SMB_01			-			++		++	++	1
SNPA	SMB_02				++					--	
SNPA	SMB_03				+		--		-	--	1
SNPA	SMB_04						-				
SNPA	SMB_05				--		+		+	++	
SNPA	SMC_01				++					++	
SNPA	SMC_02									--	
SNPA	SMC_03									++	
SNPA	SMC_05										
SNPA	SMC_06				--						
SNPA	SMD_01						-		-	+	
SNPA	SMD_10		+							-	
SNPA	SMD_12		-				+				
SNPA	SME_01									+	
SNPA	SME_02						--	-	--		1
SNPA	SME_03						-				
SNPA	SMF_01										
SNPA	SMF_02		+		++		++	++	++	++	1
SNPA	SMF_03									--	
SNPA	SMF_04									++	
SNPA	SMF_05										
SNPA	SMG_01							-			
SNPA	SMG_02				-			--			
SNPA	SMG_03			-						-	
SNPA	SMG_04									--	
SNPA	SMG_05		-					++		--	1
SNPA	SMG_08										
SNPA	SMG_09									--	
SNPA	SMG_10									--	
SNPA	SMH_01				+					-	
SNPA	SMH_02				-					+	
SNPA	SMH_03									--	
SNPA	SMH_04										
SNPA	SMH_05				-					++	
SNPA	SMH_07									++	

Dominio	Reactivo	SEXO				NIVEL SOCIOECONÓMICO					
		DIF Uniforme	DIF No Uniforme			SELEC	DIF Uniforme	DIF No Uniforme			SELEC
			1	2	3			1	2	3	
FEM	SMA_03										
FEM	SMA_04			-					++		
FEM	SMA_06								+		
FEM	SMA_07			++			+		--		
FEM	SMB_06			-			-				
FEM	SMB_07								++		
FEM	SMB_08								--		
FEM	SMB_09			+		++	+	++		1	
FEM	SMB_10								++		
FEM	SMD_02										
FEM	SMD_03			--		+	+		++		
FEM	SMD_04					--		--		1	
FEM	SMD_05					--		--	--	1	
FEM	SMD_06					+	--	+	--		
FEM	SMD_07						+				
FEM	SMD_08			--					++		
FEM	SMD_09					+			--		
FEM	SME_04						--				
FEM	SME_05			+			+		++		
FEM	SME_06			+					--		
FEM	SME_07										
FEM	SME_08										
FEM	SMF_06						-		--		
FEM	SMF_07			+					--		
FEM	SMF_08						--				
FEM	SMF_10			+					--		
FEM	SMG_06								--		
FEM	SMG_07					++		++	++	1	
FEM	SMH_08			-					++		
FEM	SMH_09								--		
FEM	SMH_10			+			-		++		
FEM	SMH_11					--	--	--	--	1	
MI	SMA_09								-		
MI	SMA_10										
MI	SMA_11								-		
MI	SMB_11			-		++	+	++	++	1	
MI	SMB_12			+		-	--		-		
MI	SMB_13										
MI	SMC_04							+			
MI	SMC_07			-					+		
MI	SMC_08		+			++	++	++	++	1	
MI	SMC_09								-		
MI	SMC_10					--		-		1	

Dominio	Reactivo	SEXO					NIVEL SOCIOECONÓMICO					
		DIF Uniforme	DIF No Uniforme			SELEC	DIF Uniforme	DIF No Uniforme			SELEC	
			1	2	3			1	2	3		
MI	SMC_11											
MI	SMC_12						-					
MI	SMD_11					--		-			1	
MI	SME_09					--	-	-			1	
MI	SME_10				-					++		
MI	SME_11				-	++	++	++	++	++	1	
MI	SME_12					++		++	--		1	
MI	SME_13											
MI	SMF_09			-						+		
MI	SMF_11						-	-				
MI	SMF_12				--							
MI	SMF_13		+				-			+		
MI	SMG_11											
MI	SMG_12											
MI	SMH_06			-	+							
MI	SMH_12											
MI	SMH_13						+			++		
<b>TOTAL SELECCIONADOS</b>						0	<b>TOTAL SELECCIONADOS</b>					18
+ DIF moderado a favor de Mujeres ++ DIF enorme a favor de Mujeres - DIF moderado a favor de Hombres -- DIF enorme a favor de Hombres						+ DIF moderado a favor de NSE Alto ++ DIF enorme a favor de NSE Alto - DIF moderado a favor de NSE Bajo -- DIF enorme a favor de NSE Bajo						

## Anexo 21. Análisis de sesgo del reactivo SMB-01

El ítem SMB-01 (ver Figura 31) pertenece al eje de la asignatura de Matemáticas llamado sentido numérico y pensamiento algebraico; en el programa de Matemáticas se ubica dentro del tema significado y uso de las operaciones y el subtema problemas aditivos. En el Excale, este reactivo pertenece a la especificación denominada resolver problemas aditivos que impliquen el uso de números con signo.

En el poblado donde vivo se registró una temperatura de  $12^{\circ}\text{C}$  el domingo a mediodía; durante las siguientes 18 horas la temperatura descendió  $15^{\circ}\text{C}$ . ¿Qué temperatura marcó el termómetro después de las 18 horas transcurridas?

$-3^{\circ}\text{C}$  \*

$3^{\circ}\text{C}$

$27^{\circ}\text{C}$

$-27^{\circ}\text{C}$

\* Opción correcta

**Figura 31. Reactivo SMB-01**

En la primera etapa de la investigación, el SMB-01 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.98 lógitos con respecto a los de NSE bajo, lo que implica que de haber sesgo, habrá algunos elementos que aún no correspondiendo a lo que el reactivo pretende evaluar, obstaculizan para que estudiantes de NSE bajo y con la habilidad matemática suficiente puedan responderlo de manera correcta.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 4.

De los seis alumnos entrevistados, cuatro de ellos respondieron al ítem de manera correcta, y los dos que no lo hicieron provinieron de la modalidad de telesecundaria con NSE bajo. Por su parte los tres docentes lo resolvieron de manera correcta así como también utilizaron un procedimiento de solución de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que de los cuatro alumnos que lo respondieron de manera correcta, tres de ellos emplearon un procedimiento adecuado, tal como realizar una resta directa o con apoyo de una recta numérica donde coloquen los grados centígrados iniciales (12° C) menos los grados que descendió la temperatura después de 18 horas (15° C); asimismo, otro alumno (de NSE medio) aunque seleccionó la opción correcta, aparentemente adivinó, como se explicará más adelante.

Por su parte, los dos alumnos de NSE bajo que no respondieron de forma correcta el SMB-01, utilizaron procedimientos incorrectos, tales como realizar una suma, una división o multiplicar las temperaturas inicial y final (ver Tabla 23).

**Tabla 23. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMB-01, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 2)	Nivel Socioeconómico	Bajo	2	0	0	--
			Medio	0	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 4)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	1	0	1
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>3</b>
	<b>Total</b>			<b>2</b>	<b>1</b>	<b>0</b>	<b>3</b>

**Fuente:** elaboración propia

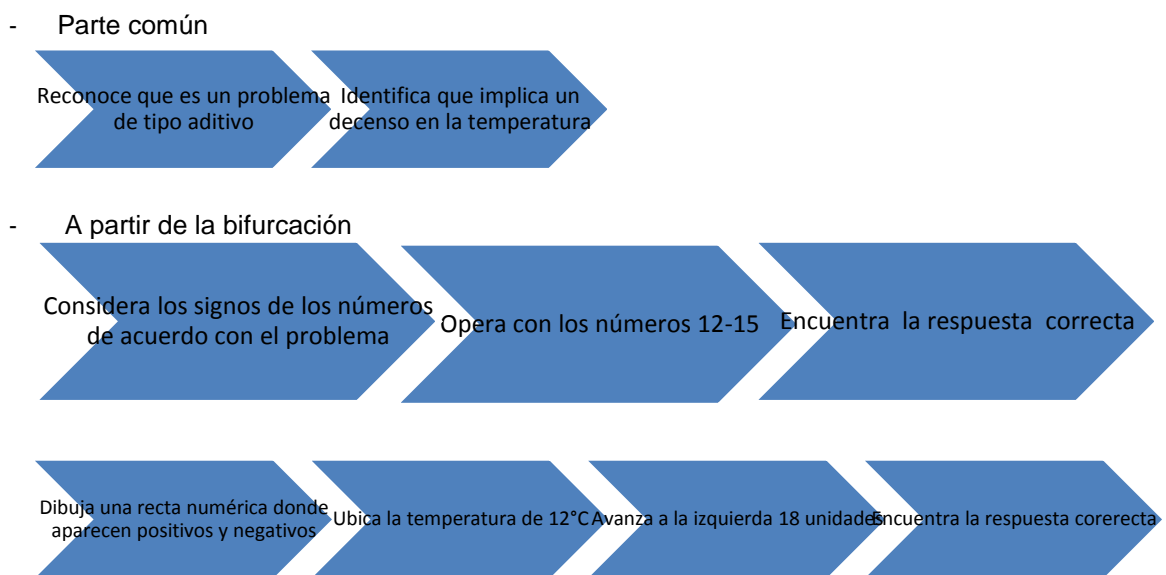
En los análisis de los protocolos verbales de las técnicas de pensamiento en voz alta y sondeo concurrente se exploró la posible presencia de cuatro problemas de varianza irrelevante al constructo del reactivo SMB-01, todos ellos en correspondencia con los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; 3) Procedimientos e instrucciones complejas, poco claras e intuitivas; y 4) Incomprensibilidad o problemas de legibilidad del ítem (Thompson *et al.*, 2002).

### ***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Sumar y restar números naturales, fraccionarios y decimales positivos o negativos.
- Conocer la secuencia numérica ascendente y descendente a partir de cualquier número.
- Utilizar criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma o con una resta.
- Calcular sumas y restas apoyándose en la recta numérica.
- Resolver problemas aditivos de cambio y de composición.

Además, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes emplearían el siguiente proceso cognitivo, que en un inicio tiene una parte común y posteriormente los entrevistados podrían optar por alguno de los dos procesos señalados a partir de la bifurcación (ver Figura 32):



**Figura 32. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-01**

Un primer análisis para identificar el potencial problema de definición imprecisa del constructo consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el

constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían reconocer que el reactivo les planteaba un problema aditivo que implicaba un descenso de temperatura y que para operar con los valores de las temperaturas debería considerar los signos de los números, eso es coincidente con lo que pretende medir el ítem, si el sustentante es capaz de: sumar y restar números naturales, fraccionarios y decimales positivos o negativos; utilizar criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma o con una resta; calcular sumas y restas apoyándose en la recta numérica; o bien, resolver problemas aditivos de cambio y de composición.

Un segundo análisis, a partir de los reportes verbales, sugiere que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con algunas variantes, las operaciones cognitivas anticipadas por las especialistas. En otras palabras, la evidencia disponible muestra que cuando los estudiantes responden al reactivo tienen que: reconocer que el reactivo implica un problema de tipo aditivo con números con signo; conocer la secuencia numérica ascendente y descendente a partir de cualquier número; utilizar criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma o con una resta; y calcular la resta (de manera opcional apoyándose en la recta numérica).

Por ejemplo, cuando la entrevistadora le pregunta a uno de los alumnos que respondió de manera correcta el reactivo si es claro lo que tiene que hacer, responde:

“sí...[ ] lo que el problema básicamente dice es que a las 12 del domingo estaba a 12 grados, durante las 18 horas descendió 15 [grados], entonces ¿qué temperatura marcó el termómetro después de las 18 horas transcurridas? Es decir, 12 grados menos 15 grados... no está tan difícil” (E07).

Otro de los alumnos que lo respondió correctamente, verbaliza mientras lee el pie del ítem “es una simple resta...” (E05). En ninguno de los entrevistados se encontraron evidencias de que al resolver el SMB-01 evocaron conocimientos o habilidades distintos a los que el reactivo pretende medir, por tanto, el reactivo no parecer presentar el problema de una definición imprecisa del constructo (ver Tabla 24).

### ***Ítem no accesible o sesgado***

Cuatro de los entrevistados (3 que respondieron correctamente el ítem y 1 que no seleccionó la respuesta correcta), aportaron elementos que apuntan que el reactivo SMB-05 no guarda uno de los principios fundamentales del Diseño Universal: la accesibilidad e imparcialidad (ver Tabla 24).

Para el grupo de NSE bajo, la utilización de ciertos términos en el pie del ítem puede obstaculizar que los sustentantes muestren sus habilidades matemáticas a cabalidad. El proceso empleado por uno de los estudiantes de TVS es muy ilustrativo de cómo el desconocimiento de la palabra “descendió” ocasionó que se eligiera una opción incorrecta. Este alumno, después de leer el pie del ítem, reconoce que se trata de un problema de adición y resuelve la operación  $12+15=27$ , luego selecciona la opción correspondiente a  $27^{\circ}\text{C}$ , que es errónea.

Este alumno no consideró de manera correcta el signo de las temperaturas, porque para el entrevistado la palabra “descendió” significó un aumento en la temperatura; cuando verbalizaba su comprensión de la tarea evaluativa, dijo:

“Aquí observo que aumentó tres grados durante esas 18 horas” (E01).

De hecho, el entrevistado señala que como la temperatura inicial es de  $12^{\circ}\text{C}$  y luego aumenta  $15^{\circ}\text{C}$ , las opciones de respuesta con números negativos no pueden ser las correctas; esto se ilustra cuando dijo:

“...menos tres grados centígrados no puede ser, si la temperatura está aumentando...” (E01).

Su comprensión errónea de la palabra “descendió” se pudo corroborar en otras partes de la entrevista, por ejemplo, cuando explica cómo descartó algunas de las opciones de respuesta señala que si el reactivo plantease “un incremento proporcional de  $3^{\circ}\text{C}$  por cada hora”, algunos alumnos también podrían elegir la opción de  $3^{\circ}\text{C}$ , sin embargo, el entrevistado señala que eso no es lo que plantea el problema, y por lo cual no seleccionó la segunda opción.

La evidencia obtenida a través del pensamiento en voz alta, también permitió conocer que el estudiante disponía de los conocimientos previos señalados en la especificación del reactivo, tales como suma de números naturales, utilizaba criterios, en forma intuitiva, para determinar si una situación problemática se resuelve con una suma, y resolución de problemas aditivos de cambio y de composición; sin embargo, la inclusión en el reactivo de una palabra que es poco familiar para el estudiante, y la cual es clave para la comprensión cabal del problema a resolver, tal como lo señaló una de las docentes, y del procedimiento matemático que debe realizar, se convirtió en un obstáculo para que el alumno escogiera la opción correcta.



El desconocimiento o comprensión de una palabra, tal como “descendió”, para ciertos grupos de población, denota que el reactivo si bien mide aspectos sobre el conocimiento matemático, también mezcla cuestiones de vocabulario, que corresponde más bien a otro de los subdominios que evalúa el Excale 09 para la asignatura de Español: comprensión lectora o eventualmente reflexión sobre la lengua. Y esta situación no es deseable.

Por otro lado, el reactivo al parecer también puede ser respondido correctamente a partir de la información del contexto de ciertos subgrupos poblacionales. Una de las alumnas (de NSE medio) que lo respondió de forma correcta, pero que no utilizó algún procedimiento matemático para llegar a la solución, utilizó su conocimiento empírico sobre las temperaturas durante el día y la noche, para acotar las opciones sobre las cuales adivinó la respuesta. Cuando se le preguntó a la estudiante porque seleccionó la opción A (la correcta), respondió:

“cambió... al ser de día, va oscureciendo, va oscureciendo y pues va bajando más la temperatura, porque en el día hace más calor y cuando va anocheciendo baja la temperatura y hace más frío...” (E04).

Si se toma en cuenta que durante su proceso para la resolución del problema no realizó ninguna resta y además, ella dudaba entre las dos opciones que tenían números negativos ( $-3^{\circ}$  y  $-27^{\circ}$ ), se puede inferir que ella dedujo la respuesta correcta, a partir de su conocimiento del medio. La alumna verbalizó:

“yo creo que es ésta [refiriéndose a las opción a)]... pero también podría ser ésta [señalando la opción d)]” (E04).

### ***Procedimientos e instrucciones complejas, poco claras e intuitivas***

Este ítem, al igual que los 18 reactivos que contenía el cuadernillo se responden seleccionando una de cuatro posibles opciones de respuesta. Cuando los entrevistados respondieron los reactivos no se identificó que tuvieran dificultades para saber cómo tenían que registrar sus respuestas, que consistía en rellenar o marcar el alveolo situado al lado izquierdo de cada una de las opciones, como lo ilustra la Figura 33 en la que se recupera una sección de uno de los reactivos en el cuadernillo ya respondido por un estudiante de NSE bajo.

¿Cuál de las siguientes afirmaciones describe lo que ocurre con las rectas k, l, m, n con respecto de la recta p?

- Tienen una pendiente m igual que la de p, ya que pasan por un mismo punto.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m menor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es menor.

**Figura 33. Recorte de un ítem respondido por uno de los alumnos de NSE bajo**

Otros de los entrevistados en lugar de rellenar los alveolos, los tachaba (ver Figura 34). Eso sería un problema en una aplicación de gran escala, donde se utilizan hojas preparadas para lectura óptica; sin embargo, en el marco de este estudio cognitivo eso no se consideró como evidencia de que las instrucciones fueran complejas poco claras e intuitivas, puesto que la primera de las instrucciones del cuadernillo señalaba: “Responde cada una de las siguientes preguntas, marcando el óvalo correspondiente para seleccionar la respuesta correcta. Si necesitas hacer operaciones, las puedes realizar en el espacio disponible en la página...”; es decir, se le pedía que *marcara* el óvalo correspondiente y no necesariamente que los rellenara. Además, al cuestionarles a los entrevistados la claridad sobre las instrucciones para responder el cuadernillo, todos manifestaron que les resultaron claras.

-3°C  
 3°C  
 27°C  
 -27°C

12 hrs - 12°C  
18 hrs - 15°C  
-----  
-03°C

**Figura 34. Recorte de un ítem respondido por uno de los alumnos de NSE medio**

En ninguno de los 18 ítems que se analizaron mediante los protocolos verbales se encontraron evidencias que indicaran la presencia de procedimientos e

instrucciones complejas, poco claras e intuitivas. Por tal motivo, ya no se presentan los resultados de este análisis para los reactivos subsiguientes.

***Incomprensibilidad o problemas de legibilidad del ítem.***

Tres de los entrevistados (un estudiante de cada NSE) aportaron evidencias de que el reactivo SMB-01 tiene problemas de legibilidad (ver Tabla 24). Aunque ninguno de los tres alumnos tuvo problema para pronunciar alguna de las palabras, dos de ellos tuvieron que leer el reactivo tres y cuatro veces, respectivamente, para proceder a responderlo.

Al parecer, la sintaxis del pie del reactivo ocasiona problemas de fluidez en su lectura. Así, cuando dos de los alumnos que tuvieron que leer el pie del ítem en repetidas ocasiones, detenían su lectura antes de “el domingo a mediodía”.

Además, cuando a uno de los estudiantes (de NSE alto) que respondió correctamente el reactivo se le preguntó si tuvo alguna dificultad para comprender lo que el ítem le planteaba, señaló que no fue propiamente una dificultad, pero que al ir leyéndolo, específicamente en la última parte que dice “¿Qué temperatura marcó el termómetro después de las 18 horas transcurridas?”, pensó que el ítem sería más difícil y que se le preguntaría por la temperatura para una hora en específico, lo que implicaría primero calcular el descenso de temperatura por hora y posteriormente calcular la temperatura final (por ejemplo, después de 18 horas) con respecto a la temperatura inicial. Eso aporta elementos para advertir una complejidad innecesaria en la redacción del pie del ítem; probablemente no se requiere que al final se diga “después de las 18 horas transcurridas”.

**Tabla 24. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMB-01, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo				
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem	
¿El entrevistado respondió correctamente el ítem?	No (N =2)	NSE	Bajo	0	1	0	1	
			Medio	0	0	0	0	
			Alto	0	0	0	0	
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	
	Sí (N =7)	NSE	Bajo	0	1	0	0	
			Medio	0	2	0	1	
			Alto	0	0	0	1	
			<b>Subtotal</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>2</b>	
	<b>Total</b>				<b>0</b>	<b>4</b>	<b>0</b>	<b>3</b>
	<b>Notas:</b> NSE=Nivel Socioeconómico <b>Fuente:</b> elaboración propia							

## Anexo 22. Análisis de sesgo del reactivo SMB-03

El ítem SMB-03 (ver Figura 35) pertenece al eje de la asignatura de Matemáticas llamado sentido numérico y pensamiento algebraico; en el programa de Matemáticas se ubica dentro del tema significado y uso de las operaciones y el subtema Potenciación y radicación. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen el uso de leyes de exponentes.

Diana resolvió la siguiente expresión:

$$\frac{(4^2)(4)(4^4)(4^0)}{(4^3)(4^5)} = \frac{1}{4^4}$$

La maestra de Matemáticas dijo a Diana que el resultado que obtuvo era incorrecto. ¿Cuál es la solución que Diana debía obtener?

- 0
- 4
- 1
- $\frac{1}{4}$  \*

\* Opción correcta

**Figura 35. Reactivo SMB-03**

En la primera etapa de este trabajo de investigación, el SMB-03 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.81 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.81 lógitos más fácil para aquellos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico alto.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 5.

De los seis estudiantes que respondieron el reactivo, dos de ellos (de NSE bajo) lo respondieron de forma incorrecta y los cuatro alumnos restantes (de NSE medio y

alto) seleccionaron la opción correcta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que de los cuatro alumnos que respondieron el ítem de forma correcta, sólo uno de ellos utilizó algún procedimiento de solución adecuado, en el que empleó las leyes de los exponentes para realizar la suma o resta de exponentes, según la operación requerida para resolver la expresión. Este alumno conocía con precisión aspectos de la ley de los exponentes que resultaban esenciales para responder correctamente al reactivo, tales como: a) en una multiplicación de potencias con la misma base, los exponentes se suman; b) en una división de potencias con la misma base, los exponentes se restan; y c), en una potencia con exponente negativo ( $a^{-n}$ ) su término equivalente es igual al valor inverso con exponente positivo ( $1/a^n$ ).

Otro de los alumnos que respondió de manera correcta, hubo algunos elementos del proceso que desconocía tal como las leyes de los exponentes, sin embargo, logró responder correctamente el reactivo porque las expresiones con exponentes las convirtió a números naturales, operó con esas cantidades y llegó al resultado correcto. Asimismo, los dos entrevistados de NSE medio que seleccionaron la opción correcta, aparentemente adivinaron pues durante las entrevistas verbalizaban que no sabían qué operación emplear para resolver la expresión; uno de ellos inicia sumando tanto las bases como los exponentes (lo que denota desconocimiento de las leyes de los exponentes); al no coincidir su resultado con ninguna de las opciones, se pregunta si lo que debió realizar es multiplicar los términos, restarlos o sumarlos.

Por su parte, de los dos alumnos que respondieron de manera incorrecta al ítem, uno de ellos en realidad no intentó solucionar el problema porque afirmó que desconocía el procedimiento que debía emplear para enfrentarse a la tarea evaluativa y dejó sin seleccionar alguna opción; el otro alumno, utilizó un procedimiento inapropiado en el que cometió muchos errores aritméticos al intentar convertir los términos con potencias a números naturales y desconocía las leyes de los exponentes (ver Tabla 25).

Vale señalar que los tres docentes utilizaron un procedimiento de solución adecuado. Para resolver el siguiente cociente de producto de potencias  $\frac{(4^2)(4)(4^4)(4^0)}{(4^3)(4^5)}$  emplearon las leyes de los exponentes para sumar los exponentes cuando los términos se tenían que multiplicar y la resta cuando se requería dividirlos.

**Tabla 25. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMB-03, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 2)	Nivel Socioeconómico	Bajo	1	0	0	--
			Medio	0	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 4)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	2	0	0
			Alto	--	0	1	1
			<b>Subtotal</b>	<b>0</b>	<b>2</b>	<b>1</b>	<b>1</b>
<b>Total</b>			<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	

**Fuente:** elaboración propia

Enseguida se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que se utilizaron en esta investigación y que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem. Aunque también se analizó el potencial problema denominado *Procedimientos e instrucciones complejas, poco claras e intuitivas*, se omite porque no se encontraron evidencias de su presencia.

***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Resolución de operaciones básicas con números enteros, decimales y fracciones.
- Potenciar números naturales y decimales.
- Razonamiento lógico matemático.

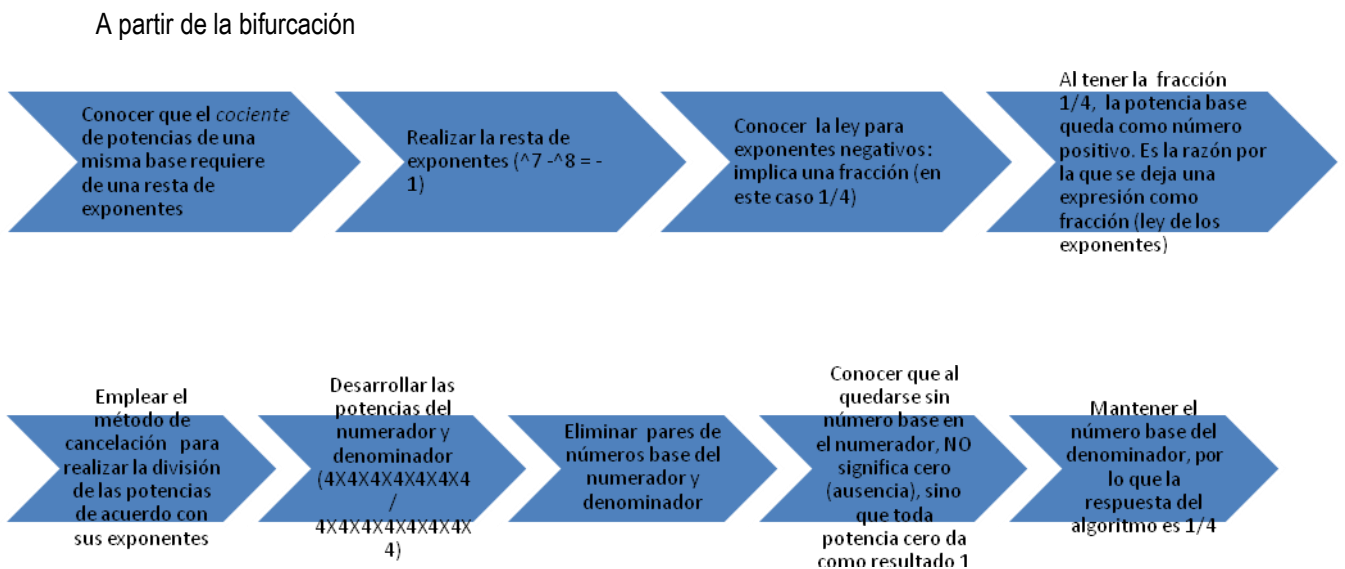
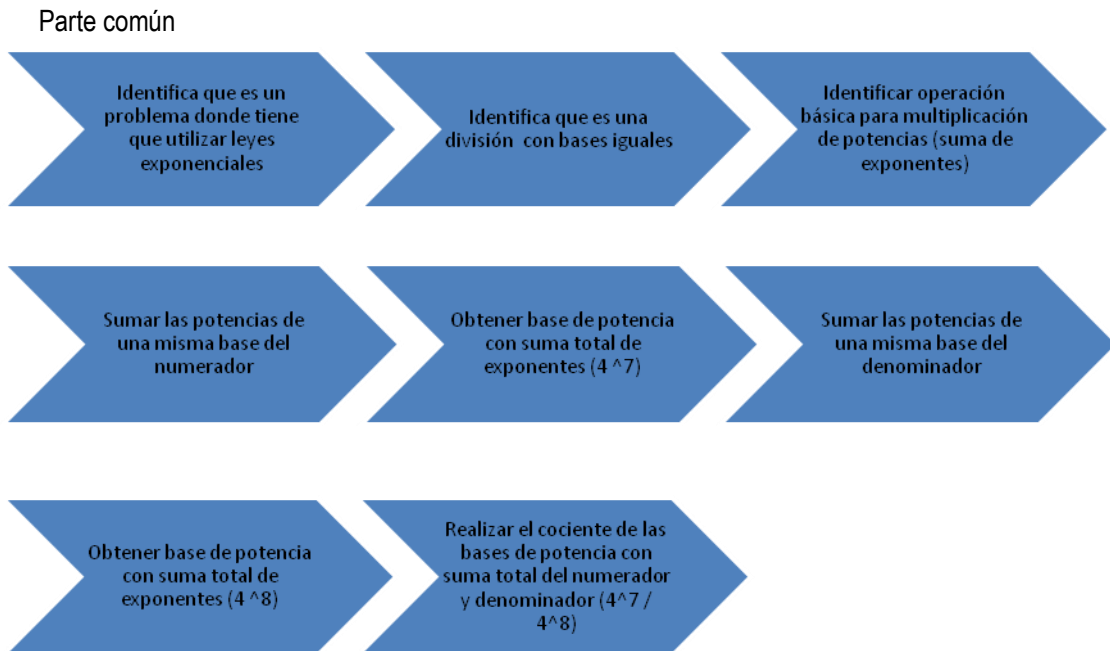
- Interpretar problemas verbales o numéricos.
- Aplicación de leyes de los exponentes.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes emplearían el siguiente proceso cognitivo, que en un inicio tiene una parte común y posteriormente los entrevistados podrían optar por alguno de los dos procesos señalados a partir de la bifurcación (ver Figura 36):

Un primer análisis para identificar el potencial problema de definición imprecisa del constructo consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían ser capaces de identificar que el problema implica la utilización de las leyes de exponentes, sumar los exponentes al multiplicar potencias de una misma base, restar los exponentes al dividir potencias de una misma base, eso es coincidente con lo que pretende medir el ítem, si el sustentante es capaz de: potenciar números naturales y decimales; y aplicar leyes de los exponentes.

Un segundo análisis, a partir de las entrevistas cognitivas, permitió se identificar tres casos, uno de cada NSE, que muestran evidencia de que el reactivo SMB-03 evoca conocimientos y habilidades distintas a las que pretende medir (ver Tabla 26). Como se establece en la especificación del reactivo, este ítem pretende medir si los sustentantes conocen y son capaces de aplicar las leyes de los exponentes; no obstante, tres alumnos que desconocían las leyes de los exponentes siguieron un procedimiento distinto mediante el cual se puede llegar la respuesta correcta.





**Figura 36. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-03**

El alumno de NSE alto que no conocía las leyes de los exponentes. Cuando se le pregunta si comprendió el problema, responde:

“sí, pero no sé qué tengo que hacer... no sé si éstos [refiriéndose a los términos con exponentes] los tengo que multiplicar... la verdad no sé qué hacer” (E08).

Pese a ello, este alumno pudo resolver correctamente el reactivo siguiendo un método que implica convertir los exponentes en números naturales. Así, mediante el desarrollo de la siguiente expresión llegó al resultado correcto sin aplicar alguna de las leyes de los exponentes, que es lo que el ítem pretende medir:

$$\frac{(4^2)(4)(4^4)(4^0)}{(4^3)(4^5)} = \frac{(8)(4)(128)(1)}{(32)(512)} = \frac{(8)(512)}{16384} = \frac{4096}{16384} = 0.25 = \frac{1}{4}$$

Los otros dos entrevistados que también aportaron evidencias sobre una definición imprecisa del constructo, pese a que uno de ellos no respondió correctamente el reactivo y el otro al final adivinó y acertó la respuesta correcta, siguieron el mismo procedimiento que el alumno anterior; sin embargo, al desarrollar las expresiones cometieron varios errores aritméticos. De haber realizado correctamente las operaciones aritméticas para convertir de números exponenciales a números naturales, muy posiblemente hayan respondido de manera correcta, lo cual implica que el ítem puede ser resuelto empleando habilidades y evocando conocimientos diferentes a los planteados por el constructo que se desea medir, que en este caso son las leyes de los exponentes.

### ***Ítem no accesible o sesgado***

No se encontró evidencia de que el ítem presente algún sesgo o sea inaccesible para alguno de los tres grupos de alumnos que se exploraron en este trabajo (ver Tabla 26).

Sin embargo, tres de los entrevistados (de NSE bajo y medio) señalan que el trabajo con expresiones con potencias es escaso y las leyes de los exponentes es un tema aislado en el currículum, que se revisan una sola ocasión y no se vuelven a trabajar más con él. La siguiente cita de uno de los docentes lo ilustra:

...la gran desventaja de esto es que... se ven estos temas y luego no se vuelven a ver nuevamente en otros bloques o en otras secuencias de aprendizaje... a veces pareciera que están un poco... que los contenidos están un tanto aislados, y como que no tienen cierta continuidad... (E03).

Por su parte, una de las alumnas señaló que ella no está acostumbrada a trabajar con expresiones como las que aparecen el reactivo. Afirma que ella con las

expresiones que está más familiarizada es con las que “no llevan los números arriba” (E04), refiriéndose a los exponentes. Otro de los alumnos también aportó elementos en el sentido de las leyes de los exponentes es un tema aislado en el currículo pues para justificar por qué no sabía cómo resolver este ítem, el estudiante señala que el tema de exponentes lo revisaron cuando estuvo en segundo grado y que ya no lo recuerda:

“nada más que esto lo vimos en segundo... no recuerdo exactamente cómo era lo de las potencias... [ ] ...no recuerdo si se sumaban cuando era multiplicación o si se restaban... me falta conocimiento sobre eso, porque lo vimos a principios de segundo y ya no lo recuerdo...” (E01).

Aunque la muestra de estudiantes no pretende ser estadísticamente representativa, ocurrió que sólo uno de los alumnos utilizó las leyes de los exponentes para resolver correctamente el reactivo. Acompañado por las expresiones de uno de los docentes y dos alumnos, bien valdría la pena considerar la relevancia de incluir en las pruebas de logro con referencia al currículo, una temática que, de ser cierto lo que afirma el profesor, tiene pocos temas que dependen o se nutren de él (lo que en términos de reticulación se denomina contenidos aislados, a diferencia de los contenidos rama o de enlace, que dan y reciben servicios simultáneamente).

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem (ver Tabla 26). Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su desconocimiento de las leyes de los exponentes, o bien, a que cometían errores aritméticos cuando emplearon algún método de resolución que implicaba convertir las expresiones con potencias a números naturales.

Durante las entrevistas, ninguno de los entrevistados mostró evidencias de que hubo una comprensión inadecuada del pie del ítem o que tuviera problemas de legibilidad. Incluso los alumnos que respondieron de manera incorrecta el ítem, verbalizan una comprensión apropiada de la tarea evaluativa, que se ilustra en la siguiente conversación entre uno de ellos y la entrevistadora:

- E01: “según esto, esto está mal: un cuarto a la cuarta potencia”
- Entrevistadora: “¿por qué... cómo sacaste esa conclusión?”

- E01: “por el problema, que me está diciendo que el resultado que obtuvo era incorrecto”
- Entrevistadora: “entonces, ¿la redacción de la pregunta te parece clara?”
- E01: “Sí”

**Tabla 26 Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMB-03, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 2)	NSE	Bajo	1	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 7)	NSE	Bajo	0	0	0	0
			Medio	1	0	0	0
			Alto	1	0	0	0
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Total</b>				<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

### Anexo 23. Análisis de sesgo del reactivo SMB-09

El ítem SMB-09 (ver Figura 37) pertenece al eje de la asignatura de Matemáticas llamado Forma, espacio y medida; en el programa de Matemáticas corresponde al tema Formas geométricas y al subtema Cuerpos geométricos. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de identificar desarrollos planos de cubos, prismas o pirámides.

Una fábrica necesita cajas cúbicas para empaclar galletas. ¿Cuál de los siguientes desarrollos planos corresponde al modelo de la caja que se necesita?

\*

\* Opción correcta

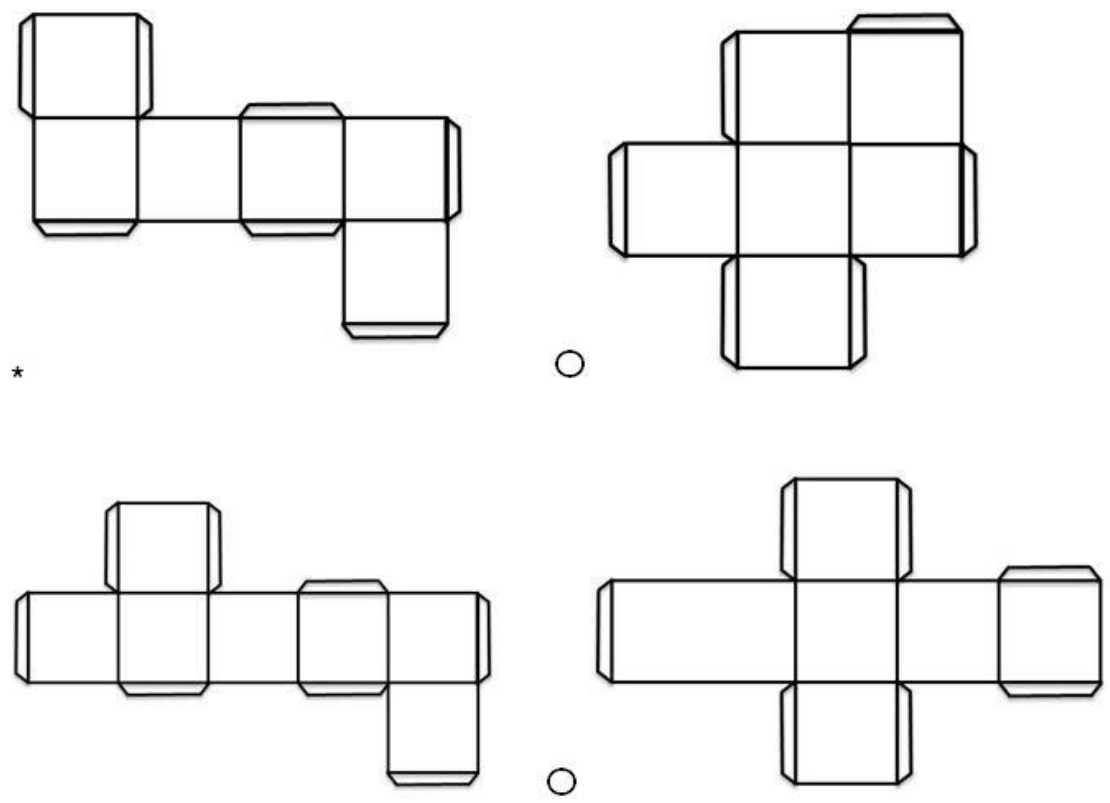
The image shows four different nets of a cube, each with a radio button next to it. The top-left net is a 1-4-1 configuration with the top square on the left, four squares in the middle, and the bottom square on the right. The top-right net is a 1-3-2 configuration with one square on top, three in the middle, and two on the bottom. The bottom-left net is a 2-3-1 configuration with two squares on top, three in the middle, and one on the bottom. The bottom-right net is a 1-3-2 configuration with one square on top, three in the middle, and two on the bottom, but with a different arrangement of flaps than the top-right net. The first radio button is marked with an asterisk, and the text '\* Opción correcta' is located below the bottom-left net.

Figura 37. Reactivo SMB-09

En la primera etapa de este trabajo de investigación, el SMB-09 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.66 lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.66 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 6.

De los seis estudiantes que respondieron el reactivo, dos de ellos (de NSE medio) lo respondieron de forma incorrecta y los cuatro alumnos restantes (de NSE bajo y alto) seleccionaron la opción correcta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que los cuatro alumnos y los tres docentes que respondieron de forma correcta el reactivo utilizaron un proceso de solución apropiado. Dichos procesos podrían ser más formales, reconociendo las características de los cubos (6 caras del mismo tamaño); o bien, a partir de la imaginación espacial de cómo se armaría cada cuerpo a partir de los cuatro desarrollos planos ofrecidos como opciones de respuesta. Tres de los alumnos que respondieron correctamente al ítem utilizaron su imaginación espacial y el otro estudiante además siguió un proceso a partir del reconocimiento de las características de los cubos. Por su parte, de los tres docentes resolvieron correctamente el ítem a partir de la identificación de las características de los cubos y uno de ellos además de basarse en el proceso más formal, también utilizó su imaginación espacial y apoyándose en dibujos que él mismo realizó, trató de armar el cubo a partir de algunos de los desarrollos planos que se incluían en las opciones de respuesta.

De los dos alumnos que no seleccionaron la opción correcta, uno de ellos desconocía algunos elementos del proceso de solución, porque trató de modificar uno de los desarrollos planos que se le ofrecieron como opción de respuesta para que se pudiera armar un cubo a partir de él. El otro alumno, utilizó un procedimiento incorrecto para solucionar el reactivo pues si bien seleccionó la opción de respuesta que contenía un desarrollo plano con 6 caras, no consideró que una de esas caras es mayor que el resto, y el cubo está construido por 6 caras del mismo tamaño (ver Tabla 27).

**Tabla 27. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMB-09, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 2)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	1	0	1	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
	Sí (N = 4)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>
	<b>Total</b>			<b>1</b>	<b>0</b>	<b>1</b>	<b>4</b>

**Fuente:** elaboración propia

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

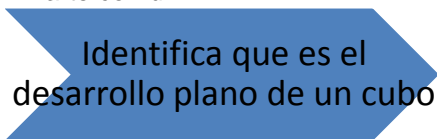
***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

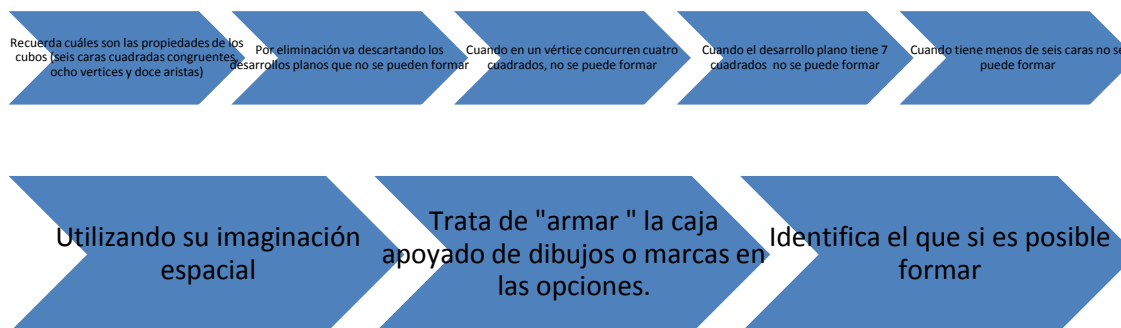
- Trazo de líneas paralelas, líneas perpendiculares, triángulos, cuadrados, rectángulos, pentágonos, hexágonos, entre otros.
- Identificar caras, vértices, aristas, ángulo diedro, poliedro, diagonales, superficies, volúmenes, por mencionar algunos.
- Determinar dónde deben ir las uniones y las pestañas en el desarrollo plano de cubos, prismas y pirámides.
- Trazar los desarrollos planos de cubos, prismas y pirámides.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes emplearían el siguiente proceso cognitivo, que en un inicio tiene una parte común y posteriormente los entrevistados podrían optar por alguno de los dos procesos señalados a partir de la bifurcación (ver Figura 38):

- Parte común



- A partir de la bifurcación



**Figura 38. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-09**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas, se encontró coherencia entre ambas fuentes.

Asimismo, un segundo análisis a partir de los reportes verbales, sugiere que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con algunas variantes, las operaciones cognitivas anticipadas por las especialistas (ver Tabla 28). Es decir, la evidencia disponible muestra que cuando los estudiantes responden al reactivo tienen que: identificar que el ítem implica el desarrollo plano de un cubo, y a partir de ahí el proceso puede seguir una ruta más formal donde se utilizan las propiedades de los cubos para descartar las opciones incorrectas, o bien, se utiliza una ruta menos formal a partir de la imaginación espacial del desarrollo plano.

Al preguntarle a uno de los alumnos que respondió correctamente al reactivo cómo comprendió el reactivo o qué es lo que la tarea evaluativa le pedía, responde:



“...me pedía, nada más... habla de una fábrica que hace galletas... cajas... las quiere empaquetar entonces necesita el modelo para hacer una caja...” (E08).

La comparación entre los reportes verbales con los procesos cognitivos anticipados por las especialistas y a su vez la contrastación con la especificación del reactivo apunta que este ítem no presenta el problema de sesgo sobre una definición imprecisa del constructo.

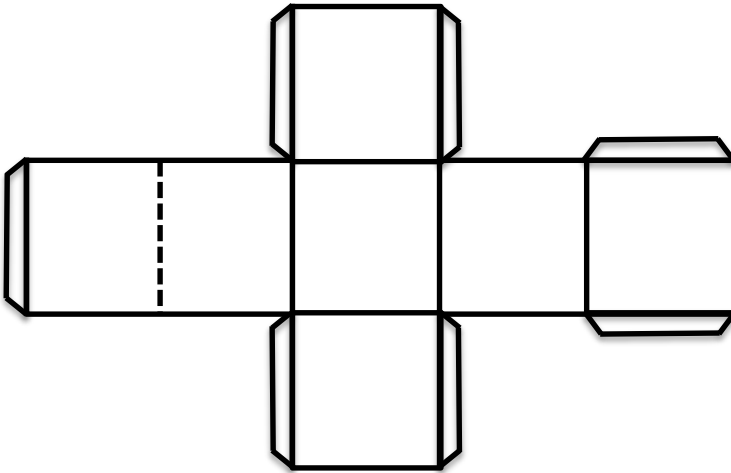
### ***Ítem no accesible o sesgado***

Una de las especialistas, señaló que el reactivo podría tener algún sesgo cultural, pues “al iniciar con la frase ‘Una fábrica necesita’, puede distraer a niños de contextos indígenas ya que no es un concepto cercano a ellos, además se está personificando a la fábrica. Los niños de contextos indígenas probablemente nunca han visto una fábrica, aunque sí sepan de su existencia por medio de imágenes, sin embargo, planteada así la pregunta considero que les crea una distracción no necesaria”; sin embargo, reconociendo que este trabajo no incluyó a estudiantes del medio indígena, la evidencia a partir de las técnicas de pensamiento en voz alta y sondeo concurrente no permite afirmar que el reactivo SMB-09 tenga algún sesgo de los alumnos de NSE bajo, pues justamente los alumnos de este estrato respondieron correctamente el ítem, al igual que los de NSE alto (ver Tabla 28). Y los alumnos de NSE medio que fueron los que respondieron el ítem de forma incorrecta, uno de ellos tuvo dificultades para comprender lo que significó “caja cúbica” mientras que otro seleccionó una respuesta incorrecta por un posible problema de ilegibilidad del ítem, que será señalada en el apartado correspondiente.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

A partir del análisis de las técnicas de pensamiento en voz alta y de los sondeos concurrentes, se encontró una posible falla del reactivo en su diseño, que provocó problemas de legibilidad con una de las opciones de respuesta (ver Tabla 28). Una alumna de NSE medio que respondió de manera incorrecta el reactivo y el docente de escuela particular, cuando resolvían el reactivo, hicieron modificaciones a uno de los planos que se ofrecen como distractores en la opción d), de manera que una de sus caras, al agregarle una arista, le permitiese formar una caja cúbica en la que una de sus caras llevaría doble grosor, como se muestra en la línea punteada de la Figura 39. Si bien el reactivo incluye como opción de respuesta un desarrollo plano en el que se formaría una caja cúbica sin desperdiciar material (que es la opción correcta), el ítem no advierte que los

desarrollos planos ofrecidos en las opciones no se pueden modificar, y por tanto, aunque no es eficiente que una de las caras de la caja tenga el doble de grosor, bajo la lógica de que los planos aceptan modificaciones, con las opciones c) y d) también podrían armarse cajas cúbicas.



**Figura 39. Modificación realizada a la opción d) del reactivo SMB-09**

**Tabla 28. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMB-09, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

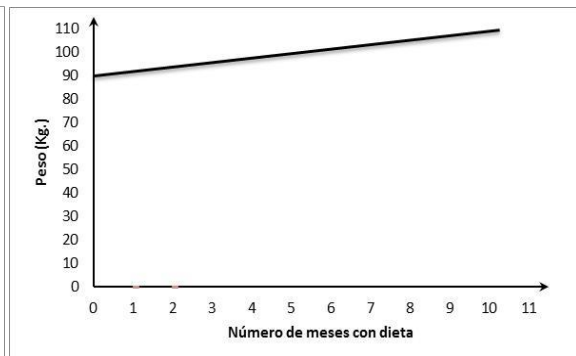
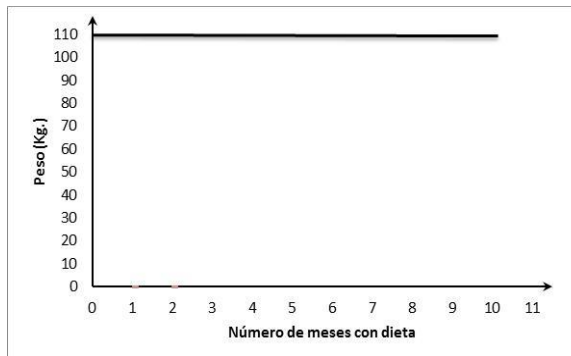
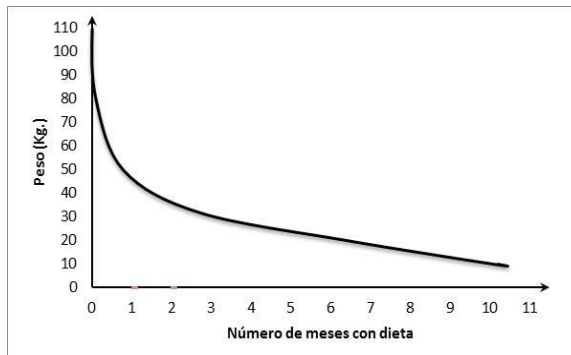
				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 2)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	1
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
	Sí (N = 7)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	1
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
<b>Total</b>				<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

## Anexo 24. Análisis de sesgo del reactivo SMB-11

El ítem SMB-11 (ver Figura 40) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Gráficas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen calcular la razón de cambio de funciones lineales.

Don Bernardino ingresó a un programa para controlar su peso durante 10 meses, al empezar pesaba 110 kg. y cada mes bajó 2 kg. ¿Cuál gráfica describe esta situación?



\* Opción correcta

Figura 40. Reactivo SMB-11

En la primera etapa de este trabajo de investigación, el SMB-11 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.69 lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.69 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 7.

De los seis estudiantes que respondieron el reactivo, sólo uno de ellos (de NSE medio) lo respondió de forma incorrecta y los cinco alumnos restantes (de los tres NSE) seleccionaron la opción correcta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que los cinco alumnos y los tres docentes que respondieron de forma correcta el reactivo utilizaron un proceso de solución apropiado que implicó identificar que el problema involucra una relación proporcional con un decremento constante, asimismo, en las gráficas que se ofrecen como opciones de respuesta tener que identificar las variables que se incluyen en cada eje (peso y meses del programa para controlar el peso), identificar aquella gráfica que cumpla con las siguientes tres condiciones: ilustrar que el peso inicial de Don Bernardino es 110 kg., presentar una reducción mensual de 2 kg durante los 10 meses que estuvo bajo un programa para controlar su peso e ilustrar el peso final de 90 kg.

El alumno que seleccionó una opción incorrecta (a), utilizó un procedimiento de solución incorrecto. Si bien advirtió que la gráfica debería ilustrar un decremento, no comprendió que tal descenso debería ser constante. A pesar que pudo descartar correctamente las opciones cuyas gráficas representaban un valor constante (opción c) y un aumento de peso (opción d), no logró identificar que la opción a) que eligió si bien representaba un descenso del peso no era constante. Su argumento para seleccionar esa gráfica fue que esa opción ilustra cómo al inicio Don Bernardino pesaba 110 kg. y la dieta que siguió por 10 meses; lo que evidencia que este alumno no supo distinguir que el punto (10,10) del plano cartesiano representaba que don Bernardino pesaba 10 kg. después de 10 meses de estar bajo el programa de control de peso, situación que no es plausible y difiere de lo que solicitaba la tarea evaluativa (ver Tabla 29).

**Tabla 29. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMB-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 1)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	1	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 5)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	1
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>5</b>
	<b>Total</b>			<b>1</b>	<b>0</b>	<b>0</b>	<b>5</b>

**Fuente:** elaboración propia

En los párrafos siguientes se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis utilizadas en esta investigación que surgen a partir de tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

### ***Definición imprecisa del constructo***

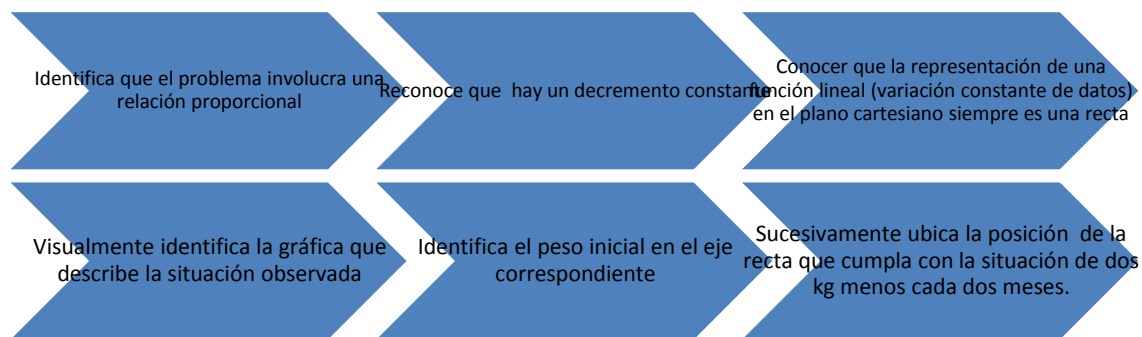
Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Números enteros, fraccionarios y decimales.
- Recta numérica.
- Plano cartesiano.
- Gráficas en el plano cartesiano.
- Operar con números enteros, fraccionarios y decimales (al hacer operaciones del cálculo de la razón de cambio).
- Ubicar correctamente puntos en la recta numérica (al identificar la escala de los ejes).

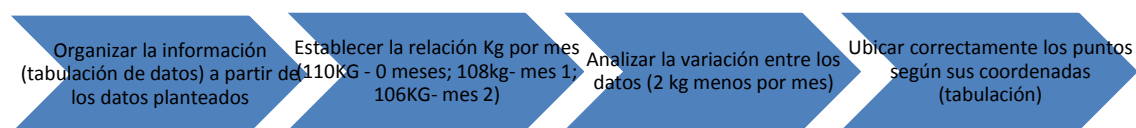
- Localizar puntos en la recta numérica dadas sus coordenadas y viceversa, dado un punto identificar las coordenadas (al identificar los puntos que forman parte de la gráfica de la situación que se está trabajando).
- Interpretar gráficas en un plano cartesiano (saber leer la gráfica para establecer la razón de cambio que está representada, en el caso de una curva, para identificar que se trata de una variación o razón de cambio que no es constante).

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes podrían emplear alguno de los dos siguientes procesos cognitivos, que pueden ir paralelo (ver Figura 41):

Proceso 1:



Proceso 2:



**Figura 41. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMB-11**

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían identificar que el problema involucra una relación proporcional con un decremento constante, conocer que la representación de una función lineal (variación constante de datos)

en el plano cartesiano siempre es una recta, identificar el peso inicial en el eje correspondiente y ubicar la recta que cumpla con la situación de dos kg menos cada dos meses, eso es coincidente con lo que pretende medir el ítem, si el sustentante conoce: el plano cartesiano, las gráficas en el plano cartesiano y, además si es capaz de: interpretar gráficas en un plano cartesiano (saber leer la gráfica para establecer la razón de cambio que está representada, en el caso de una curva, para identificar que se trata de una variación o razón de cambio que no es constante) .

Un segundo análisis, a partir de las entrevistas cognitivas, permitió confirmar que el proceso cognitivo que emplearon los estudiantes para enfrentarse a la tarea evaluativa es consistente con los conocimientos y habilidades que se plantean en la especificación del reactivo y en el proceso previsto por las especialistas en enseñanza de las matemáticas, es decir, no se encontró evidencia que indique el ítem tiene una definición imprecisa del constructo que desea evaluar.

Los entrevistados identificaron que el problema involucra una relación proporcional, identificaron que la gráfica debería ilustrar un decremento constante, conocían que la representación de una función lineal en el plano cartesiano siempre es una recta, identificaron el peso inicial de 110 kg. en el eje de las Y, y ubicaron la posición de la recta que cumpliera con la situación de dos kg menos cada dos meses.

Para responder el reactivo, los cinco alumnos que lo resolvieron de manera correcta utilizaron la estrategia de ir descartando opciones. Así, todos descartaron en primera instancia la gráfica que ilustrada un valor constante, luego descartaron la gráfica que mostraba un comportamiento ascendente y, finalmente, descartaron la gráfica con un función curvilínea porque advirtieron que el descenso de peso en el primer mes fue muy acelerado (60 kg.), lo cual no correspondía, señalaron, con lo planteaba la tarea evaluativa.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMB-11 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE; en resumen, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***



El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Un docente (de telesecundaria) señaló que el reactivo podría tener una dificultad innecesaria al no incluir una cuadrícula dentro de cada gráfica. Las líneas secundarias del plano cartesiano, servirían para que los sustentantes pudieran identificar de manera más directa si el comportamiento mensual descrito en el pie del reactivo, se cumple en las gráficas. Sin embargo, no se consideró un problema porque el mismo profesor lo plantea como algo no indispensable para resolver correctamente el reactivo, sino como un arreglo que ayudaría a encontrar la respuesta con más facilidad. Además, ocho de los nueve entrevistados respondieron de forma correcta el ítem y no señalaron que éste fuera una dificultad para haber localizado la respuesta correcta.

## **Anexo 25. Análisis de sesgo del reactivo SMC-08**

El ítem SMC-08 (ver Figura 42) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Diagramas y Tablas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas de conteo que impliquen utilizar recursos gráficos (diagramas de árbol o tablas de doble entrada).

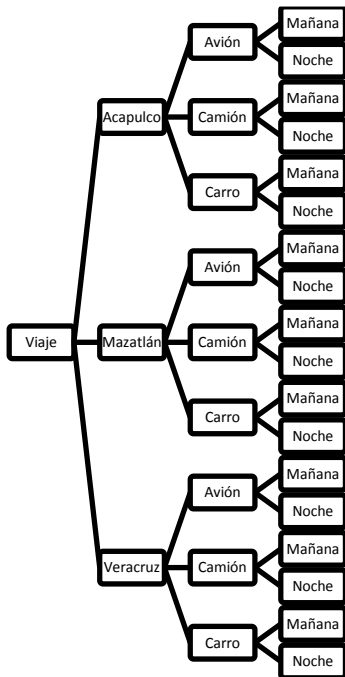
En la primera etapa de este trabajo de investigación, el SMC-08 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 1.21 lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 1.21 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 8.

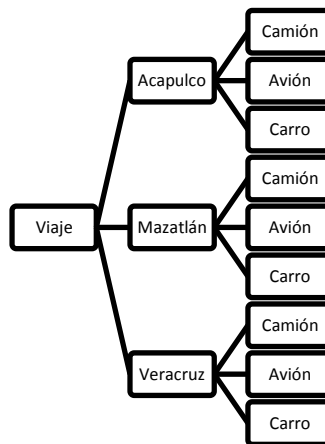
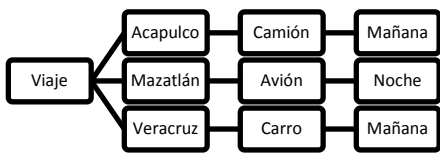
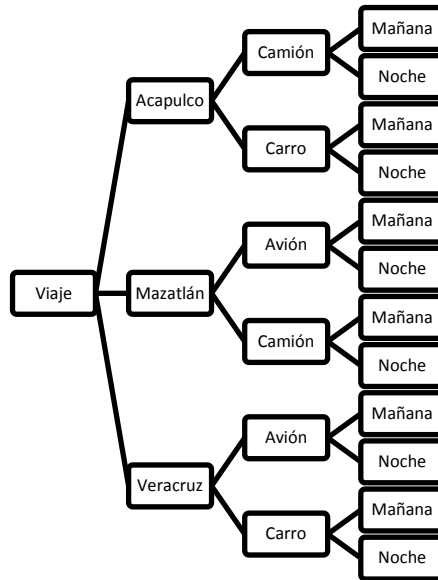
De los seis estudiantes que respondieron el reactivo, sólo uno de ellos (de NSE medio) lo respondió de forma incorrecta y los cinco alumnos restantes (de los tres NSE) seleccionaron la opción correcta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que los cinco alumnos y los tres docentes que respondieron de forma correcta el reactivo utilizaron un proceso de solución apropiado que implicó identificar que el reactivo implicaba un problema de conteo con distintas tres variables: la primera el destino del viaje podría tener tres opciones: Acapulco, Mazatlán o Veracruz; la segunda variable, el medio de transporte, que también podría tener tres valores: avión, camión o automóvil; y la tercera variable que refiere a la hora de salida la cual podría tener dos opciones: mañana o noche; y posteriormente identificar el diagrama de árbol que contenía todos los resultados posibles, es decir, aquél que representaba a las tres variables y sus respectivas opciones de respuesta de tal manera que se formara un diagrama de árbol de  $3 \times 3 \times 2 = 18$  ramas (ver Tabla 30).

Una persona realizará un viaje. Tiene la opción de ir a Acapulco, Veracruz o Mazatlán; puede hacerlo en avión, automóvil o camión, por la mañana o por la noche. ¿Cuál es el diagrama de árbol que muestra todas las opciones posibles?



\*



\* Opción correcta

Figura 42. Reactivo SMC-08

La alumna que seleccionó una opción incorrecta, de NSE medio, utilizó un procedimiento de solución incorrecto. Esta entrevistada, seleccionó el diagrama de la opción b) el cual en la variable de medio de transporte no incluye alguno de sus tres posibles valores. Cuando se le pidió que explicase por qué eligió esa opción, respondió:

“porque está un poquito más explicada y sin tanto espacio... sin ocupar tantos renglones... y aquí [señalando la opción a) que es la correcta] te pone tooodo, lo repite aquí [señalando la opción a)]...”  
(E04).

**Tabla 30. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMC-08, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 5)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	1	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 1)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	1
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>5</b>
<b>Total</b>			<b>1</b>	<b>0</b>	<b>0</b>	<b>5</b>	

**Fuente:** elaboración propia

Los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente, se presentan enseguida desglosándolos según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

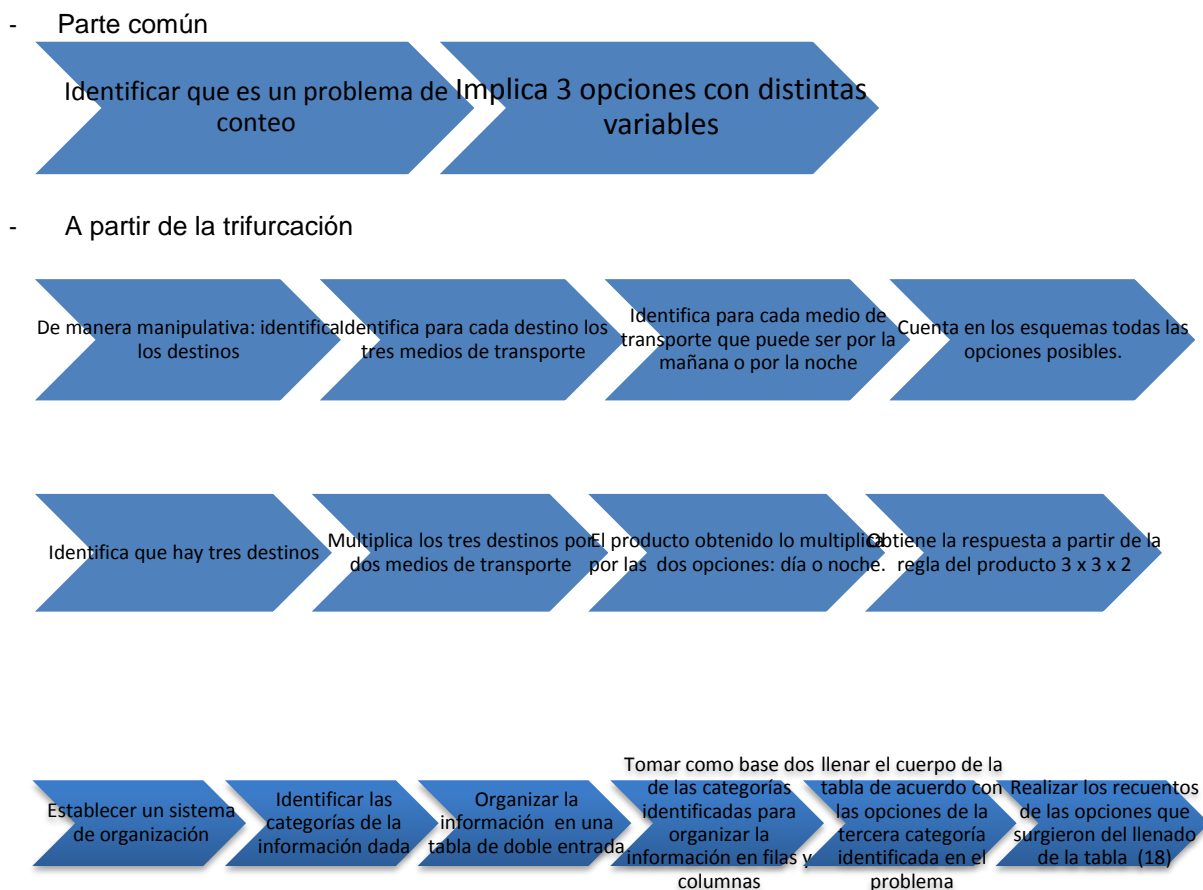
***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Interpretar información presentada en diagramas de árbol o tablas de doble entrada.
- Enumerar las formas posibles de resolver el problema.
- Representar el problema en el diagrama de árbol o tabla de doble entrada.
- Identificar el diagrama de árbol o tabla de doble entrada que contiene todos los resultados posibles.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes podrían emplear alguno de los tres siguientes procesos cognitivos, que tienen una parte común y luego se trifurcan (ver Figura 43):

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes debería identificar que es un problema de conteo con tres distintas variables: destino de viaje, medio de transporte y hora de salida, y que el esquema debería representar las tres opciones de viaje, los tres medios de transporte y las dos horas de salida; eso es coincidente con lo que pretende medir el ítem, si el sustentante es capaz de: interpretar información presentada en diagramas de árbol o tablas de doble entrada e identificar el diagrama de árbol o tabla de doble entrada que contiene todos los resultados posibles.



**Figura 43. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMC-08**

Un segundo análisis, a partir de los reportes verbales, permitió corroborar que el proceso cognitivo que emplearon los estudiantes para enfrentarse a la tarea evaluativa es consistente con los conocimientos y habilidades que se plantean en la especificación del reactivo y en el proceso previsto por las especialistas en enseñanza de las matemáticas, es decir, no se encontró evidencia que indique el ítem tiene una definición imprecisa del constructo que desea evaluar (ver Tabla 31).

Los entrevistados identificaron que el ítem planteaba un problema de conteo con tres variables distintas y luego cinco de ellos que respondieron de forma correcta siguieron un proceso donde de manera manipulativa identificaron cuál diagrama de árbol incluía los tres destinos, para cada destino los tres medios de transporte, y para cada medio de transporte los dos horarios de salida.

Asimismo, los cinco alumnos que respondieron correctamente el reactivo, en su proceso de resolución fueron descartando las opciones incorrectas cuando identificaban que cierto diagrama de árbol no incluía una variable (opción d), una de las opciones de respuesta de una de las variables (opción b), o bien, no incluía opciones de respuesta en más de una variable (opción c).

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMC-08 así como los sondeos que se hicieron durante las entrevistas cognitivas, permitieron encontrar una posible fuente de sesgo en este reactivo.

Dos de los entrevistados (un alumno de NSE alto y una docente que atiende estudiantes de NSE medio) si bien respondieron de forma correcta el ítem, advirtieron una posible confusión para cierto grupo de alumnos (ver Tabla 31).

El alumno se percata que el término “automóvil” que se utiliza en el pie del reactivo como una de los medios de transporte, en el diagrama se modificó y aparece como “carro”, y eso vuelve un poco más difícil su resolución,

“no sé si lo hicieron a propósito... creo que sí porque no suelen fallar en esas cosas [refiriéndose a las pruebas estandarizadas en gran escala]... es de que te ponen otros tipos de nombres, un sinónimo, pero es para que pienses más, supongo... es como un detalle extra para hacerlo un poco más difícil, casi nada, pero un poco” (E07).

Por su parte, la docente que atiende alumnos de NSE medio, tuvo que establecer la relación automóvil=carro, y advertía que aquellos alumnos que por su contexto no puedan establecer esa relación, es muy probable que cause confusión cuál opción elegir.

En los análisis DIF este reactivo es más difícil para los alumnos de NSE bajo. Si bien en este estudio no se encontró ese patrón, una de las especialistas que analizó el reactivo para anticipar el proceso cognitivo que emplearían los alumnos al enfrentarse al ítem, también señaló que ciertos términos podrían sesgar el reactivo, “Para alumnos de contextos indígenas los referentes de transporte pueden ser un distractor ya que muchos niños indígenas que no han salido de sus comunidades probablemente no tengan muy claro la diferencia entre automóvil y camión, y como en el enunciado se menciona ‘automóvil o camión’ pueden considerarlo como una u otra opción, por lo que el resultado puede ser incorrecto”. Esta advertencia, si bien está ligada a una clasificación de las poblaciones según su lengua materna, no obstante, en México existe una alta asociación entre NSE

bajo y lengua indígena, por lo cual es pertinente para esta investigación que clasificó a las poblaciones de estudiantes según su NSE.

**Incomprensibilidad o problemas de legibilidad del ítem.**

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem (ver Tabla 31). De hecho, como ya se señaló, ocho de los nueve entrevistados lo respondieron de forma correcta.

Por otro lado, el reactivo presenta un problema: en el pie del ítem la variable que versa sobre el horario de salida al viaje tiene dos opciones “por la mañana o por la noche”; en cambio en los diagramas de árbol que se incluyen como opciones de respuesta en lugar de “noche” se escribió “tarde”. Este error ocurrió cuando se copió el reactivo de la versión original a formato de Word, para ajustar el cuadernillo y que se pudiera aplicar sólo con los 18 ítems de las entrevistas cognitivas. Este error es atribuible al tesista y sólo ocurrió en el marco de este estudio, en el reactivo original y que se aplicó a nivel nacional, este error no está presente, por lo cual no se mencionó como posible problema, y además, cuando se analizó la presencia de alguna de las fuentes de varianza irrelevante al constructo, este error se aisló de las dificultades exhibidas por los entrevistados.

**Tabla 31. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMC-08, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 5)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 4)	NSE	Bajo	0	0	0	0
			Medio	0	1	0	0
			Alto	0	1	0	0
			<b>Subtotal</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>
<b>Total</b>			<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>	
<p><b>Notas:</b> NSE=Nivel Socioeconómico  <b>Fuente:</b> elaboración propia</p>							



## **Anexo 26. Análisis de sesgo del reactivo SMC-10**

El ítem SMC-10 (ver E-Anexo 9) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Gráficas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen la interpretación de información representada en gráficas de línea.

En la primera etapa de este trabajo de investigación, el SMC-10 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.83 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.83 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

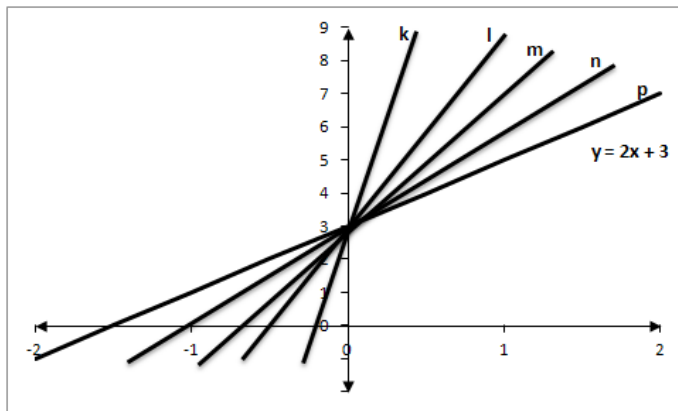
El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 9.

De los seis estudiantes que respondieron el reactivo, sólo uno de ellos (de NSE medio) lo respondió de forma correcta y los cinco alumnos restantes (de los tres NSE) seleccionaron una opción incorrecta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, el alumno que seleccionó la opción correcta del reactivo aparentemente adivinó; pues durante su proceso de resolución del problema, además de que afirmaba no estar seguro de cuál opción elegir, fue preseleccionando cada una de las cuatro opciones, para lo cual utilizaba argumentos que denotaban una incomprensión del problema y del procedimiento apropiado a utilizar, cuestión que se detallará párrafos más adelante. Por su parte, cuatro de los cinco alumnos que respondieron de manera incorrecta, también utilizaron un procedimiento incorrecto para enfrentarse a la tarea evaluativa pues en su gran mayoría desconocían el concepto de pendiente, o bien, la interpretación adecuada de incremento del ángulo de inclinación, cuestiones que resultaban indispensables para resolver el reactivo. Un alumno (de NSE alto) que tampoco respondió de forma correcta el ítem, si bien comprendía el concepto de pendiente, hubo algunos elementos del proceso que desconocía, tales como el sentido en que se considera un incremento del ángulo de inclinación así como que dentro de la ecuaciones de la forma  $f(x) = mx + b$ , la “m” representa la pendiente (ver E-Anexo 9).

Considera las siguientes gráficas de la forma  $f(x) = mx + b$ .

La ecuación de la recta p es  $y = 2x + 3$ .



¿Cuál de las siguientes afirmaciones describe lo que ocurre con las rectas k, l, m, n con respecto de la recta p?

- Tienen una pendiente m igual que la de p, ya que pasan por un mismo punto.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es mayor.\*
- Tienen una pendiente m menor que la de p, ya que su ángulo de inclinación es mayor.
- Tienen una pendiente m mayor que la de p, ya que su ángulo de inclinación es menor.

\* Opción correcta

**Figura 44. Reactivo SMC-10**

Por su parte, los tres docentes cuando respondieron el reactivo SMC-10 utilizaron un procedimiento de solución de manera correcta, en el que identificaron la recta que se utilizaría como referente para comparar el valor de las pendientes de las rectas k, l, m y n y seleccionar la opción de respuesta que reflejase la situación de la gráfica.

**Tabla 32. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMC-10, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 5)	Nivel Socioeconómico	Bajo	2	0	0	--
			Medio	1	0	0	--
			Alto	1	0	1	--
			<b>Subtotal</b>	<b>4</b>	<b>0</b>	<b>1</b>	<b>0</b>
	Sí (N = 1)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	1	0	0
			Alto	--	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>
	<b>Total</b>			<b>4</b>	<b>1</b>	<b>1</b>	<b>0</b>

**Fuente:** elaboración propia

Enseguida se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

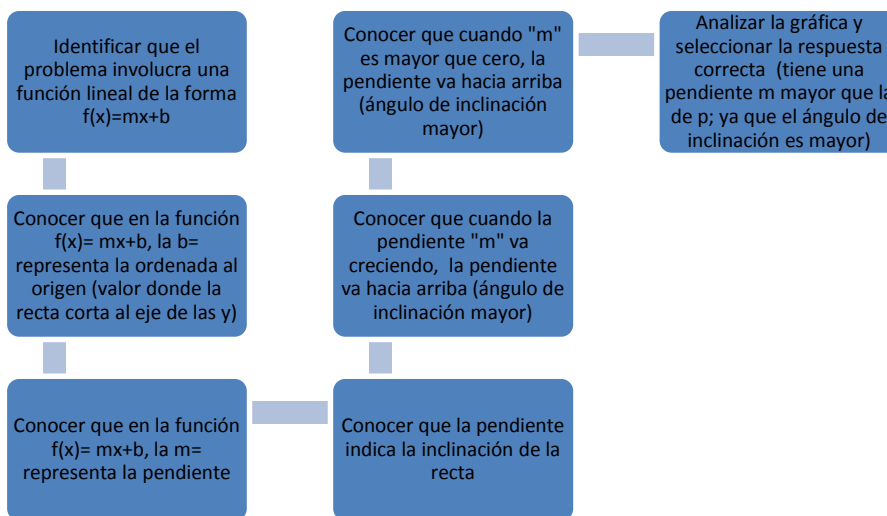
### ***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- En primer lugar, los alumnos deben comprender la información contenida en la gráfica. La comprensión de la gráfica es un proceso complejo que implica identificar claramente los elementos estructurales que, según Friel, Curcio y Bright (2001), son:
  - Contenido contextual: saber leer el título y etiquetas (para comprender qué información se muestra en la gráfica y cuáles son las variables representadas)
  - Marco del gráfico: saber identificar los ejes interpretar correctamente la escala usada en cada uno (para comprender las unidades de medida de las magnitudes representadas en el gráfico). Por lo que la interpretación

- de los números (naturales, fracciones comunes decimales) en la recta numérica es un conocimiento esencial para dar respuesta al reactivo.
- Especificadores del gráfico: saben interpretar que, en el caso de una gráfica de línea, las coordenadas de los puntos unidos por los segmentos que forman la línea, son parejas ordenadas cuyo primer elemento corresponde a la unidad del tiempo y el segundo elemento corresponde a la magnitud graficada y representada en el eje vertical. Esto implica que los alumnos estén familiarizados con la representación de puntos en un plano cartesiano.
  - Para el caso del contenido que se evaluará y como se presentará una gráfica de línea con dos magnitudes que varían con respecto al tiempo, para responder al reactivo se tendrán que reconocer los elementos estructurales mencionados en ambas magnitudes por separado y, además, relacionándolos entre sí.
  - Si se decide usar porcentajes como unidad del eje vertical, el manejo de los mismos constituye un conocimiento necesario para dar respuesta al reactivo.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un proceso cognitivo lineal como el que se ilustra en la Figura 45:



**Figura 45. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMC-10**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas se encontró inconsistencia entre ambas fuentes. Mientras que las especialistas anticiparon un proceso que implicaba el conocimiento de la pendiente ( $m$ ) y la

ordenada al origen (b) en la función  $f(x)=mx+b$  así como conocer el comportamiento gráfico de la pendiente cuando el ángulo de inclinación es mayor que cero; en cambio, en la especificación del reactivo si bien se señala que los alumnos deben comprender la información contenida en la gráfica, especifican que los alumnos deben: interpretar correctamente la escala usada en cada uno ejes del plano cartesiano; interpretar los números (naturales, fracciones comunes decimales) en la recta numérica; interpretar que, en el caso de una gráfica lineal, las coordenadas son parejas ordenadas, cuyo primer elemento corresponde a la unidad del tiempo y el segundo elemento corresponde a la magnitud graficada y representada en el eje vertical, entre otras habilidades.

Un segundo análisis, a partir de la información recaba mediante las entrevistas cognitivas, permitió conocer que para enfrentarse a la tarea evaluativa ocho de los entrevistados emplearon un proceso cognitivo que es distinto con los conocimientos y habilidades que se plantean en la especificación del reactivo, es decir, la evidencia apunta a una definición imprecisa del constructo que desea evaluar (ver Tabla 33).

Si bien los alumnos no respondieron de forma correcta el ítem, cuando se enfrentaron a él, intentaban evocar conocimientos como pendiente, ordenada al origen, ecuaciones de la forma  $f(x) = mx + b$  y ángulo de inclinación, cuestiones que los docentes, quienes sí respondieron el reactivo de forma correcta, evocaron de manera adecuada.

En las entrevistas, se identificaron muchos episodios que ilustran cómo los alumnos y docentes utilizan conocimientos y habilidades distintos a los señalados en la especificación del reactivo, para tratar de encontrar la respuesta correcta. Por ejemplo, cuando una de las alumnas al ir analizando las opciones de respuesta y descartar la opción b) (que es la opción correcta) afirma que esa opción no puede ser la correcta porque "...tienen una pendiente m mayor que la de p..." (E02).

O bien, cuando a otra de las alumnas se le pide que explique por qué seleccionó la opción a) que incorrecta, justifica:

"porque tienen una pendiente igual, porque aquí se observa [mientras compara la longitud de las rectas p y m], y si las juntamos, es igual, nada más que una está más inclinada que la otra... y las dos pasan por un mismo punto..." (E04).

Un elemento fundamental para responder correctamente el reactivo es el concepto de pendiente, conocimiento distinto a lo que plantea la especificación a partir del cual se diseñó el reactivo. El entrevistado que aparentemente adivinó manifiesta

su confusión para responder al ítem de la siguiente manera: “!Bueno, la palabra que no entiendo es pendiente!... tiene una pendiente  $m$ , eso no” y más adelante en la entrevista vuelve a decir “...es que no sé qué es la pendiente...” (E05).

Por su parte, los docentes, que respondieron de manera correcta el ítem, también evocaron conocimientos distintos a los previstos en la especificación del reactivo. Se ilustra cuando uno de ellos verbaliza:

“...La dificultad de este reactivo estriba en recordar el concepto de pendiente... muchas veces son conceptos teóricos que a la hora de verlos en un esquema, en un diagrama, sí se les dificulta recordar qué es la pendiente...” (E03).

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMC-10 así como los sondeos que se hicieron durante las entrevistas cognitivas, permitieron encontrar una posible fuente de sesgo en este reactivo.

Tres de los entrevistados (los dos alumnos de NSE bajo y su profesor), experimentaron o advirtieron una posible confusión al emplear el pie del reactivo dos maneras de expresar las ecuaciones, una en su terminología general tal como  $f(x)=mx+b$ , y otra particular para la recta  $p$ ,  $y=2x+3$ . Este grupo de alumnos afirman que el término  $f(x)=mx+b$  les resulta confuso. Afirman que no entienden el significado de la expresión  $f(x)=mx+b$  y uno de ellos expresa que no sabe si tiene que multiplicar  $f$  por  $X$  o qué tiene que realizar con dicha expresión; estos alumnos no comprenden que  $f(x)$  significa “función de  $x$ ” (ver Tabla 33).

Por su parte, el docente de los alumnos de NSE bajo detecta que sus alumnos podrían confundirse con la recta que servirá de referente. Él verbaliza: “...aquí el alumno se confundiría un poco... porque le nombran de dos formas la expresión...” (E03).

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Cuatro de los entrevistados (los dos alumnos y su docente de NSE medio, así como un alumno de NSE alto) aportaron elementos en ese sentido que se describe a continuación (ver Tabla 33). Se identificaron dos problemas:

1. La inclusión de dos enunciados y una ilustración previo a la pregunta a responder, ocasionó que al menos tres entrevistados no comprendieran que lo planteaba el reactivo era un comparación de las rectas  $k$ ,  $l$ ,  $m$  y  $n$ , todas en

ellas con respecto a la recta p. Además, la inclusión del segundo enunciado “La ecuación de la recta p es  $y = 2x + 3$ ” en el pie del ítem es innecesario. No se requiere para encontrar la opción correcta, y complejiza innecesariamente la redacción del reactivo.

La docente que atiende alumnos de NSE medio, que sí realizó la comparación de las rectas con respecto a la recta p, advierte que la forma en que están redactadas las opciones es confusa. La docente señala que para que sea más clara la distinción entre las opciones b) y d), convendría que se precisara con respecto a cuál recta es la que se hace la comparación:

“...para mí, le faltaría algo. Dice ‘que su ángulo de inclinación es mayor’... con respecto a la línea original... mayor a la recta p” (E06).

La maestra señaló que de no contar con esa precisión en las opciones de respuesta (que la recta de referencia es la p), los alumnos podrían tomar como punto de referencia el eje de las ordenadas (Y), y aunque ella no lo menciona, también podrían tomar como punto de referencia el eje de las abscisas (X). Esta situación refleja que si bien el reactivo sí precisa en la pregunta que la comparación se hará con respecto a la recta p, la sintaxis empleada no es del todo adecuada.

2. Nombrar una de las rectas de la gráfica como “m” también puede llevar a una confusión innecesaria con la “m” que simboliza la pendiente. A partir de ello, y en conjunto con el primer problema presentado con anterioridad, tres de los entrevistados comprendieron que debían comparar la recta m con respecto a la recta p. Uno de los alumnos de NSE alto, que respondió correctamente la mayoría de los 18 ítems del cuadernillo, cuando lee por tercera ocasión la pregunta de este reactivo y sus opciones de respuesta, dice:

“!Esto es un poco extraño!”...[] esto no me había dado cuenta... apenas ahorita lo vi... que todas se están basando en las rectas m y en la p, entonces las k, l y m creo que no te sirven porque no vienen en las respuestas... entonces las quitaría... están ahí para confundirme, por así decirlo” (E07).

Además, una de las alumnas, además de que demostró que no comprendía el concepto de pendiente (pues lo relacionó con la longitud de las líneas), comprendió que la pregunta le planteaba una comparación entre las rectas m y p, exclusivamente. En todo su proceso de respuesta así lo manifestó. Por ejemplo, cuando analiza la opción b) para saber si la elige como correcta o no, compara las rectas m y p al tiempo que dice “yo también podría decir que es ésta, porque sí tiene un ángulo de inclinación mayor...” (E04).

**Tabla 33. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMC-10, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 5)	NSE	Bajo	2	2	0	0
			Medio	1	0	0	1
			Alto	2	0	0	1
			<b>Subtotal</b>	<b>5</b>	<b>2</b>	<b>0</b>	<b>2</b>
	Sí (N = 4)	NSE	Bajo	1	1	0	0
			Medio	2	0	0	2
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>2</b>
<b>Total</b>				<b>8</b>	<b>3</b>	<b>0</b>	<b>4</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia



## Anexo 27. Análisis de sesgo del reactivo SMD-04

El ítem SMD-04 (ver Figura 46) pertenece al eje de la asignatura de Matemáticas llamado Forma, espacio y medida; en el programa de Matemáticas corresponde al tema Medida y al subtema Estimar, medir y calcular. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de realizar conversiones de medida de superficie.

<p>Uno de los países más pequeños del mundo es Mónaco, con 2 km<sup>2</sup> de superficie total. ¿A cuántas hectáreas equivale la superficie de este país?</p> <p><input type="radio"/> 20 ha</p> <p><input type="radio"/> 200 ha*</p> <p><input type="radio"/> 2 000 ha</p> <p><input type="radio"/> 20 000 ha</p> <p>* Opción correcta</p>
--

**Figura 46. Reactivo SMD-04**

En la primera etapa de este trabajo de investigación, el SMD-04 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.80 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.80 lógitos más fácil para aquellos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 10.

De los seis estudiantes que respondieron el reactivo, sólo uno de ellos (de NSE alto) lo respondió de forma correcta y los cinco alumnos restantes (de los tres NSE) seleccionaron una opción incorrecta. Por su parte, dos docentes respondieron el reactivo de forma adecuada y uno (que atiende alumnos de NSE medio) lo respondió de manera incorrecta.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, el alumno que respondió de forma correcta utilizó un procedimiento

adecuado, en el que: 1) recordó que una hectárea mide 10,000 m<sup>2</sup>; 2) dedujo que una hectárea en forma de cuadro mediría 100 mts por lado; 3) recordó que 1 km<sup>2</sup> equivale a un cuadrado donde cada uno de sus lados mide 1 km de longitud; 4) calculó que en cada km<sup>2</sup> caben cien hectáreas y por tanto en 2 kms<sup>2</sup> caben 200 hectáreas.

Por su parte, dos de los cinco alumnos que respondieron de manera incorrecta (uno de NSE medio y otro de NSE alto), utilizaron un procedimiento correcto pero hubo algunos elementos del proceso que desconocían. Es decir, estos alumnos si bien concebían que tenían que hacer una división para transformar los kms<sup>2</sup> a hectáreas o una multiplicación para hacer la transformación en el sentido inverso, no recordaban cuánto representaba una hectárea, razón por la cual no seleccionaron la respuesta correcta. Otros dos de los cinco alumnos (de NSE bajo y NSE medio) que no respondieron de manera adecuada el reactivo desconocían el procedimiento que deberían emplear para hacer la conversión de unidades de superficie, es decir, cuando intentaban hacer la conversión a hectáreas no distinguían entre medidas de distancia o lineales como metros o kilómetros y medidas de superficie o área como m<sup>2</sup> o km<sup>2</sup>; y el quinto alumno (de NSE medio) que respondió erróneamente, intentó adivinar la respuesta (ver Tabla 34).

Dos de los tres docentes que respondieron el reactivo SMD-04 utilizaron un procedimiento de solución de manera correcta, en el que recordaron que una hectárea equivale a 10,000 m<sup>2</sup> y que si tuviera la forma de un cuadrado cada lado mediría 100 mts, también recordaron que 1 km<sup>2</sup> equivale a un cuadrado que mide 1 km por lado, lo que significa que contiene 1 millón de m<sup>2</sup>, y a partir hicieron la equivalencia de que 1 km<sup>2</sup> contiene 100 hectáreas por lo que 2 km<sup>2</sup> equivalen a 200 hectáreas. Uno de los docentes, al manifestar que no estaba seguro de cuánto equivalía una hectárea, si bien utilizó un razonamiento similar al de los otros docentes, no consiguió la respuesta correcta.

**Tabla 34. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMD-04, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 5)	Nivel Socioeconómico	Bajo	1	0	1	--
			Medio	1	1	0	--
			Alto	0	0	1	--
			<b>Subtotal</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>0</b>
	Sí (N = 1)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	0	0	0
			Alto	--	0	0	1
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
	<b>Total</b>			<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>

**Fuente:** elaboración propia

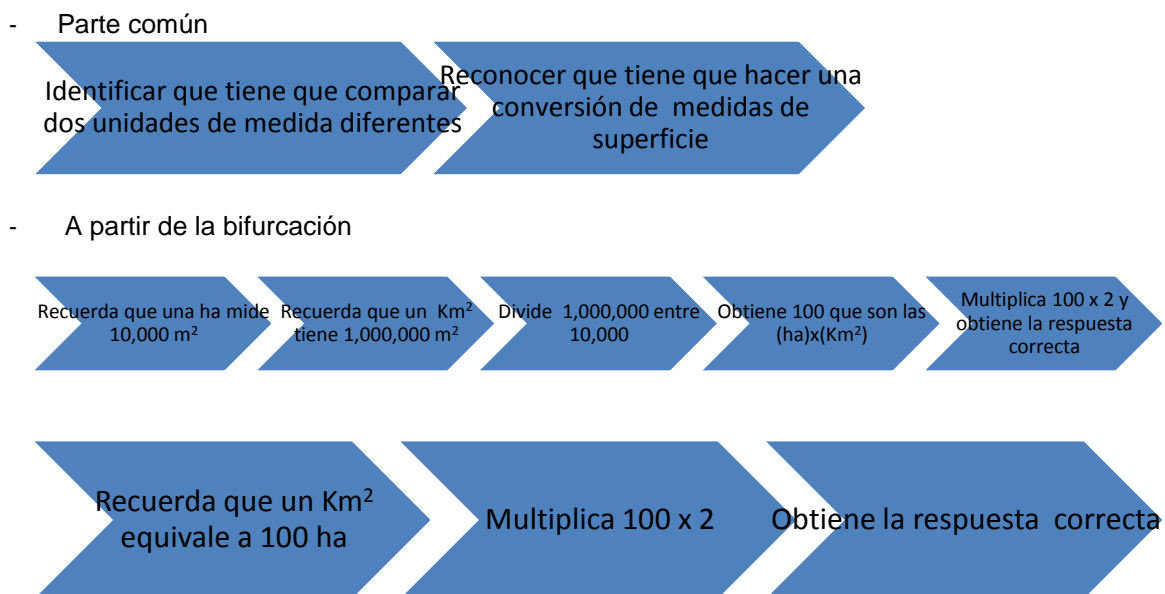
En los subapartados siguientes se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se organizaron según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones que se estudiaron en esta investigación: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Se debe tener el conocimiento de las unidades de medida de superficies, su equivalencia entre las diferentes unidades de área y la habilidad para realizar conversiones a otras unidades de área.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un proceso cognitivo que en un inicio es común y luego se bifurca, como se ilustra en la Figura 47:



**Figura 47. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMD-04**

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. El resultado de este primer análisis apunta a que este ítem no presenta el problema de definición imprecisa del constructo; cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían comparar dos unidades de medida diferentes en la que tendrían que hacer una conversión de medidas de superficie, que en este caso es de hectáreas a  $\text{km}^2$ , eso es coincidente con lo que el ítem pretende medir: si el sustentante tiene el conocimiento de unidades de medida de superficies, su equivalencia entre las diferentes unidades de área y la habilidad para realizar conversiones a otras unidades de área.

Un segundo análisis, a partir de la información recabada en los reportes verbales, sugiere que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con algunas variantes, las operaciones cognitivas anticipadas por las especialistas (ver Tabla 35). En otras palabras, la evidencia disponible muestra que cuando los estudiantes responden de forma correcta el reactivo tienen que: identificar que tienen que comparar dos medidas diferentes; que tienen que recordar a cuánto equivale una hectárea (ya sea en  $\text{m}^2$  o  $\text{km}^2$ ), o bien, calcular a cuánto equivale una hectárea a partir del recuerdo de que se compone de un área en forma de cuadrado que mide 100 mts. por lado; recordar o calcular cuántos

metros cuadrados tiene un  $\text{km}^2$ ; considerar que el problema plantea la conversión de una superficie de  $2 \text{ km}^2$ ; y realizar la transformación mediante una multiplicación cuando se convierten hectáreas a  $\text{km}^2$  o mediante una división al convertir  $\text{km}^2$  a hectáreas. Toda la evidencia apunta a que este reactivo no presenta el problema de una definición imprecisa del constructo.

Incluso los alumnos que no respondieron correctamente el reactivo, evocaron la parte fundamental del constructo que consiste en la conversión de unidades de superficie. La siguiente cita de uno de los entrevistados que no conocía la equivalencia de una hectárea lo ilustra cuando ante un sondeo que le hace el entrevistador “si supieras la equivalencia, ¿cómo lo resolverías?”, responde: “nada más transformaría estos dos kilómetros cuadrados al valor de las hectáreas” (E08).

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMD-04 así como los sondeos que se hicieron durante las entrevistas cognitivas, permitieron encontrar una posible fuente de sesgo en este reactivo.

Siete de los entrevistados (de los tres NSE) tuvieron dificultades o advirtieron que sus alumnos podrían tener dificultades para saber la equivalencia de una hectárea. El conocimiento de la equivalencia es indispensable para resolver correctamente el ítem. En este trabajo cinco de los alumnos y un docente no supieron dicha equivalencia y ese fue un factor fundamental para que no respondieran correctamente el reactivo (ver Tabla 35).

Al responder un sondeo del entrevistador sobre si le queda clara la pregunta, uno de los alumnos responde “sí, pero no me queda clara la equivalencia para convertirlo a hectáreas” (E08). Otro alumno, después de leer el pie del ítem, trata de recordar cuánto es un kilómetro y cuánto es una hectárea, y dice: “no estoy muy seguro... un hectárea serían... mmm...” y al no responder, la entrevistadora lo sondea cuestionándole si la pregunta le resulta clara, a lo que el mismo alumno responde “sí es clara, sólo que se tiene que recordar lo de kilómetros y de hectáreas” (E05).

Una de las docentes señala que la hectárea no es una medida común en su contexto escolar:

“...son medidas que no se manejan mucho. Se maneja el centímetro, se maneja el metro lineal, el kilómetro, pero la hectárea no mucho...porque una hectárea se maneja más bien, hablando de

superficies, de campo... entonces el niño no está muy familiarizado con eso...” (E06).

Además de la falta de familiaridad en el mundo real con esta medida, salvo para aquéllos que provienen de contextos rurales en los que se manejan las hectáreas para medir superficies de parcelas o ranchos, el docente que atiende alumnos de NSE bajo (y que acuden a una escuela ubicada en un contexto semi-rural) señala que el reactivo será complicado porque este tema suele estar aislado en el programa de matemáticas:

“así como está, es un tanto complejo para el alumno porque no... como le digo, es algo que no se retoma mucho a lo largo de los contenidos. Estos contenidos se abordan y ya no se vuelven a aplicar, ese es el problema, que no se aplican en otras secuencias, en otros bloques... entonces es algo que quedó aislado y que no se le dio aplicación...” (E03).

Dos de los docentes señalan que si el reactivo incluyese la equivalencia de las hectáreas, por ejemplo, en metros cuadrados, el alumno podría hacer la conversión, porque es más común para ellos saber que 1 km es igual a 1000 metros y a partir de ahí hacer las conversiones necesarias.

El reactivo SMD-04 presentó un DIF que favorece a los alumnos de NSE bajo; a partir de la información de las entrevistas cognitivas, y tomando en cuenta que existe una fuerte asociación entre NSE y ruralidad, es verosímil razonar que muy probablemente el sesgo del reactivo se derive de aquéllos alumnos que conocen la equivalencia de una hectárea, medida que suele utilizarse en medios rurales y por tanto se pueda inferir que sean los alumnos ajenos a los contextos rurales los que se vieron obstaculizados para responder correctamente el reactivo, a pesar de que podrían haber respondido un ítem similar que le solicitara convertir superficies.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem (ver Tabla 35). Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su incapacidad para realizar al menos uno de los dos siguientes procesos que resultan fundamentales para resolver el reactivo: conversión de unidades de superficie o área, o bien, conocer cuántos m<sup>2</sup> tiene una hectárea de tal manera que se pudiera realizar la conversión de 2 km<sup>2</sup> a 200 hectáreas.

**Tabla 35. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMD-04, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 6)	NSE	Bajo	0	2	0	0
			Medio	0	3	0	0
			Alto	0	1	0	0
			<b>Subtotal</b>	<b>0</b>	<b>6</b>	<b>0</b>	<b>0</b>
	Sí (N = 3)	NSE	Bajo	0	1	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>Total</b>				<b>0</b>	<b>7</b>	<b>0</b>	<b>0</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

## **Anexo 28. Análisis de sesgo del reactivo SMD-05**

El ítem SMD-05 (ver Figura 48) pertenece al eje de la asignatura de Matemáticas llamado Forma, espacio y medida; en el programa de Matemáticas corresponde al tema Medida y al subtema Estimar, medir y calcular. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de realizar conversiones de medida de superficie.

En la primera etapa de este trabajo de investigación, el SMD-05 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.66 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.66 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 11.

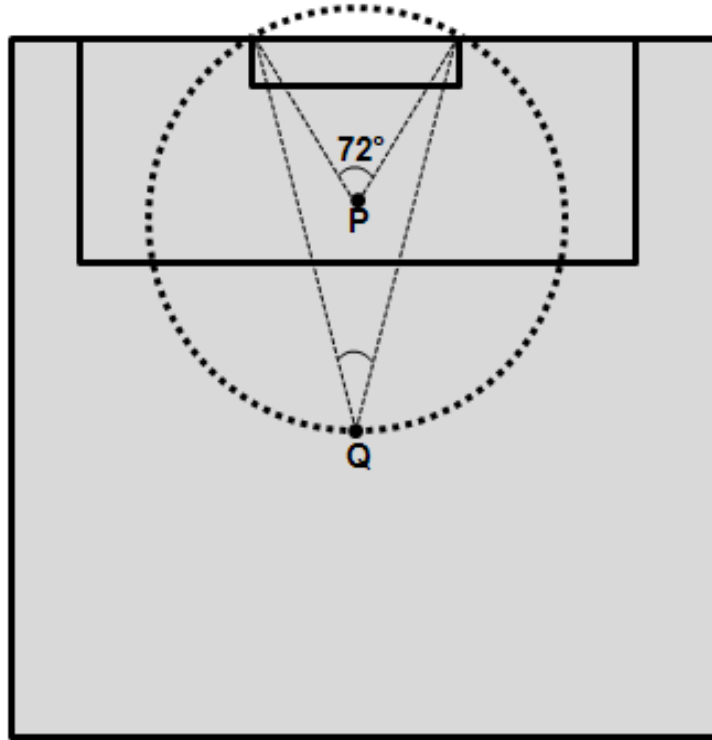
De los seis estudiantes que respondieron el reactivo, cuatro de ellos (de NSE bajo y NSE alto) lo respondió de forma correcta y los dos alumnos de NSE medio seleccionaron una opción incorrecta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los cuatro alumnos que la respondieron de forma correcta y los tres docentes utilizaron un procedimiento adecuado, en el que: identificaron que en el problema hay un ángulo central ( $P$ ) y un ángulo inscrito ( $Q$ ); identifica dichos ángulos comparten el mismo arco de la circunferencia, el cual está enmarcado por la portería; recordaron que una de las propiedades de los ángulos que tienen estas características es que el ángulo inscrito es la mitad del ángulo central; y dividieron el ángulo  $P$  entre dos.

Por su parte, los dos alumnos que no seleccionaron la opción correcta, no recordaron que cuando un ángulo central ( $P$ ) y un ángulo inscrito ( $Q$ ) comparten el mismo arco una circunferencia, el segundo es el doble del primero; por lo que utilizaron un procedimiento de solución incorrecto, tal como realizar una estimación visual del ángulo  $Q$  a partir de la figura (ver Tabla 36).



El siguiente dibujo representa a dos jugadores de futbol haciendo su práctica de rutina. El jugador P se localiza en el centro del círculo punteado, con un ángulo de tiro de  $72^\circ$  respecto a la portería. ¿Cuál es el ángulo de tiro del jugador Q?



- $18^\circ$
- $36^\circ$  \*
- $54^\circ$
- $72^\circ$

\* Opción correcta

Figura 48. Reactivo SMD-05

**Tabla 36. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMD-05, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 2)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	2	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 4)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>
<b>Total</b>			<b>2</b>	<b>0</b>	<b>0</b>	<b>4</b>	
<b>Fuente:</b> elaboración propia							

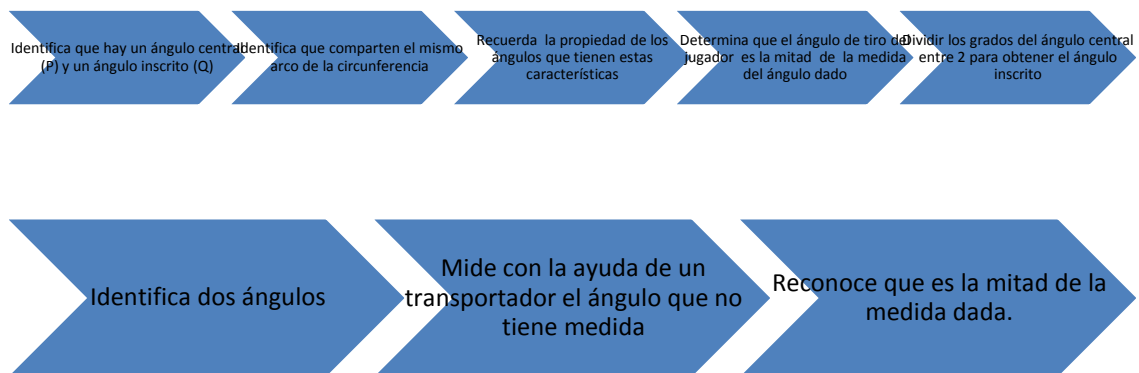
A continuación se muestran los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Se requiere el conocimiento de los conceptos de ángulo –en particular la definición de ángulo inscrito y central en una circunferencia–, diámetro, cuerda y radio de una circunferencia. También el alumno debe recordar la relación que hay entre los ángulos central e inscrito, que abarcan el mismo arco, en una circunferencia. Además se debe contar con la habilidad para hacer construcciones con regla y compás, para medir ángulos con transportador.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los dos procesos cognitivos que presentan en la Figura 49:



**Figura 49. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMD-05**

Un análisis inicial consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían: identificar que hay un ángulo central (P) y un ángulo inscrito (Q) que comparten el mismo arco de la circunferencia; recordar la propiedad de los ángulos que tienen estas características; determinar que el ángulo de tiro del jugador es la mitad de la medida del ángulo dado y; dividir los grados del ángulo central entre 2 para obtener el ángulo inscrito, lo cual es coincidente con lo que reactivo pretende medir: si el sustentante conoce los conceptos de ángulo inscrito y central en una circunferencia, así como también recordar la relación que hay entre los ángulos central e inscrito, que abarcan el mismo arco, en una circunferencia.

Un análisis más detallado, a partir de los reportes verbales, confirmó que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean las operaciones cognitivas anticipadas por las especialistas en el primer sendero cognitivo. En otras palabras, la evidencia disponible muestra que cuando los estudiantes responden de forma correcta el reactivo tienen que: conocer qué es un ángulo, una circunferencia y un arco; identificar que los ángulos de tiro de los jugadores P y Q corresponden a un ángulo central y uno inscrito en una

circunferencia que comparten el mismo arco; recordar que cuando existen esas condiciones, el valor del ángulo inscrito es la mitad del ángulo central; y operar la división para obtener el valor de Q.

Cuando a un alumno se le pide que diga con sus propias palabras cómo comprendió el ítem, su respuesta ilustra que evoca los conocimientos señalados en la especificación y que sendero cognitivo que siguió es uno de los anticipados por las especialistas:

“nada más está pidiendo... si éste jugador [señalando P en el cuadernillo] tenía un ángulo de 72 desde éste punto [volviendo a señalar P], ¿cuál es el ángulo de éste [apuntando con su lápiz el ángulo Q]?... y ya” (E08).

Uno de los docentes, después de señalar en el cuadernillo que el ángulo del jugador P se le conoce como ángulo central y el de Q se le conoce como ángulo inscrito, identifica que comparten el mismo arco de la circunferencia, cuando verbaliza: “tienen el mismo arco, que viene siendo la portería” y luego dice: “el estudiante debe recordar una regla: que uno [refiriéndose al ángulo central] es el doble del otro [refiriéndose al ángulo inscrito]” (E03). En resumen, no se encontró evidencia empírica de que este reactivo presente el problema de una definición imprecisa del constructo.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMD-05 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE. Los análisis DIF anticipaban que este reactivo es más difícil para alumnos de NSE alto, en cambio, tanto los de ese estrato como los de NSE bajo respondieron correctamente el reactivo, y los alumnos de NSE que no seleccionaron la opción correcta, su verbalización indica que eso ocurrió porque no disponían de un conocimiento que resultaba fundamental para resolver el problema: el reconocimiento de que cuando un ángulo central y uno inscrito comparten el mismo arco de una circunferencia, el primero es el doble del segundo. En resumen, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Más bien, lo que se halló es que tanto los alumnos como los

docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su incapacidad para identificar que el reactivo planteaba un problema cuya resolución implicaba reconocer la presencia de un ángulo central y uno inscrito compartiendo un mismo arco, y que la relación entre ellos está determinada por:  $\text{Ángulo Inscrito} = \text{Ángulo Central} / 2$ .

Mientras los entrevistados leyeron el pie del ítem así como cuando se les hicieron preguntas de sondeo para indagar problemas de legibilidad o incomprensión de algún término, todos afirmaron que el reactivo es claro en su formulación y que habían comprendido la tarea que tenían que realizar. Por ejemplo, al preguntarle a una docente si considera que la redacción fue adecuada así como la figura, responde: “sí...sí, también el dibujo” (E06)

## **Anexo 29. Análisis de sesgo del reactivo SMD-11**

El ítem SMD-11 (ver Figura 50) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Gráficas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de identificar el comportamiento de las gráficas lineales de la forma  $y = mx + b$  cuando el valor de  $b$  aumenta o disminuye.

En la primera etapa de este trabajo de investigación, el SMD-11 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.65 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.65 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

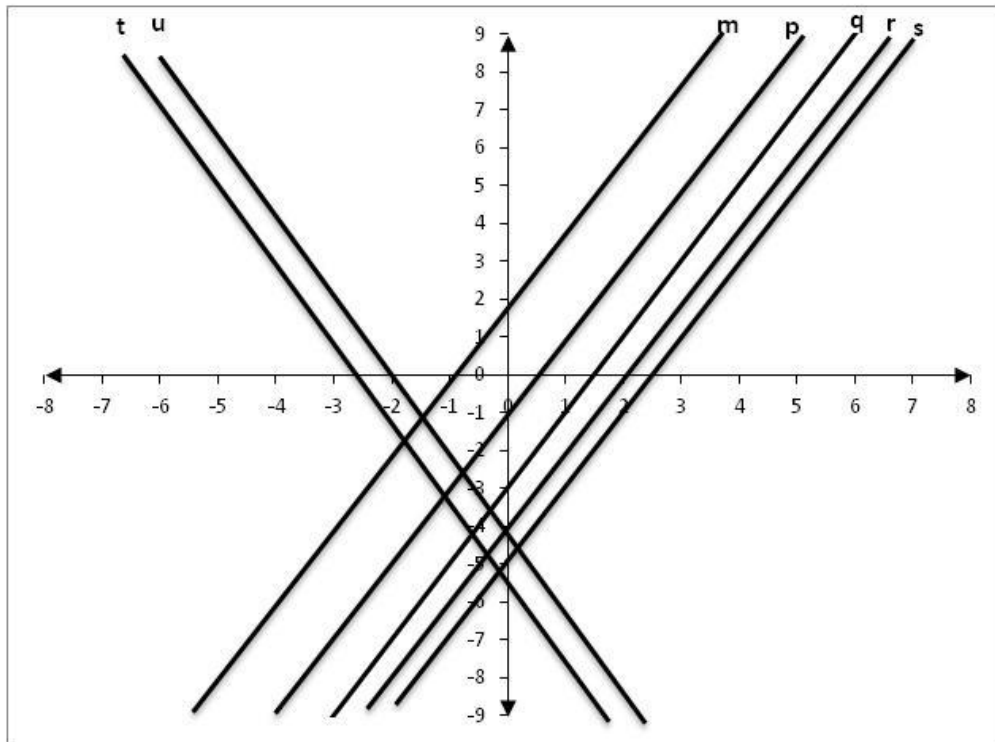
El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 12.

De los seis estudiantes que respondieron el reactivo, sólo dos de ellos (de NSE alto) lo respondieron de forma correcta y los cuatro alumnos restantes (de NSE Bajo y Medio) seleccionaron una opción incorrecta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los alumnos que respondió de forma correcta utilizaron un procedimiento adecuado, en el que: 1) identifica que el problema plantea la relación entre dos variables (función lineal) del tipo  $f(x) = mx + b$ ; 2) conoce que  $b$  es la ordenada al origen; 3) identifica cuáles rectas del plano cartesiano muestran un incremento en  $b$ , a partir de  $-3$  (ver Tabla 37).

Tres de los cuatro alumnos que respondieron el ítem de manera incorrecta (dos de NSE medio y otro de NSE bajo), desconocían el procedimiento de solución correcto a utilizar, o bien, utilizar un procedimiento incorrecto (ver Tabla 37). Por ejemplo, uno de ellos siguió un sendero cognitivo en el que trató de sustituir los valores de la fórmula  $f(x) = 2x - 3$  de la recta  $q$ , para luego tratar de identificar esos puntos en el plano cartesiano, sin embargo, no relacionó  $f(x)$  con el eje de las ordenadas, por lo cual obtuvo valores incorrectos, además de que cometió errores aritméticos al realizar la sustitución de la fórmula. Otros dos alumnos mostraron en su verbalización y procedimiento empleado que no comprendían que en la denominación general de las rectas lineales del tipo  $f(x) = mx + b$ , la “ $b$ ” representa la ordenada al origen y tampoco supieron identificar que en la recta  $q$ , el valor  $-3$  representaba el valor de  $Y$  cuando  $X$  es igual a cero.

El siguiente plano muestra rectas de funciones de la forma  $f(x) = mx + b$ , donde la ecuación de la recta q es  $f(x) = 2x - 3$ . ¿Qué rectas muestran un incremento en el valor de b?



- r y s
- t y u
- m y p\*
- s y t

\* Opción correcta

Figura 50. Reactivo SMD-11

Un alumno (de NSE bajo) dejó sin responder el reactivo, porque al leer el pie del reactivo afirmó que éste le resultó confuso, en específico en el significado de la “fórmula”  $f(x)=mx + b$ . En la entrevista cognitiva no se pudo apreciar los procesos cognitivos que empleó, porque el alumno no intentó solucionar el ítem.

Por su parte, los tres docentes utilizaron un procedimiento de solución apropiado, similar al que emplearon los dos alumnos que respondieron correctamente el ítem.

**Tabla 37. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMD-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 4)	Nivel Socioeconómico	Bajo	1	0	0	--
			Medio	2	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 2)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	0	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>
<b>Total</b>			<b>3</b>	<b>0</b>	<b>0</b>	<b>2</b>	
<b>Fuente:</b> elaboración propia							

En los apartados siguientes se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

#### ***Definición imprecisa del constructo***

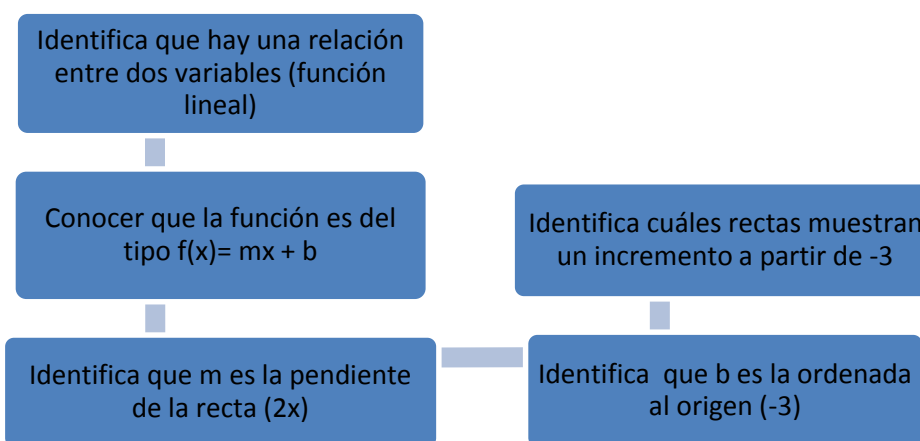
Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Relación funcional de dos variables.
- Localización de puntos en el plano cartesiano.
- Identificar elementos relevantes en una gráfica.



- Identificar el parámetro  $b$  en la intersección de la recta con el eje de las ordenadas.
- Identificar el efecto que tiene en la gráfica el incremento o decremento del parámetro  $b$ .
- Realizar la transferencia mental de un registro gráfico al algebraico.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un proceso cognitivo que es lineal, como se ilustra en la Figura 51:



**Figura 51. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMD-11**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas se encontró que coherencia entre ambas fuentes. Así, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían: identificar que el problema plantea una relación entre dos variables (función lineal) del tipo  $f(x)= mx + b$ ; identificar que  $m$  es la pendiente de la recta ( $2x$ ); identificar que  $b$  es la ordenada al origen ( $-3$ ) e; identificar cuáles rectas muestran un incremento a partir de  $-3$ ; lo cual coincide con lo que el ítem pretende medir, cuando en su especificación se prevé que será respondido correctamente cuando el sustentante sea capaz de: identificar una relación funcional de dos variables; identificar el parámetro  $b$  en la intersección de la recta con el eje de las ordenadas; identificar el efecto que tiene en la gráfica el incremento o decremento del parámetro  $b$ ; y realizar la transferencia mental de un registro gráfico al algebraico.

Un análisis más detallado, a partir de la información obtenida mediante las entrevistas cognitivas corroboró que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean las operaciones cognitivas anticipadas por los especialistas (ver Tabla 38). En otras palabras, la evidencia disponible muestra que cuando los estudiantes y docentes responden de forma correcta el reactivo tienen que: identificar la relación entre dos variables (función lineal) del tipo  $f(x) = mx + b$ ; conocer que “b” es la ordenada al origen; localizar puntos en el plano cartesiano; identificar el parámetro b en la intersección de la recta q con el eje de las ordenadas; e identificar cuáles rectas del plano cartesiano muestran un incremento en b, a partir de -3.

Por ejemplo, uno de los estudiantes de NSE alto que identifica que “b” es la ordenada al origen y que en la recta q toma un valor de -3, cuando se le pregunta qué es lo que está buscando en el plano cartesiano, responde “cual está cruzando por el -3... ésta [señalando en el plano cartesiano a la recta q] la q” (E08).

Uno de los docentes, evidencia que evoca que “b” es la ordenada al origen cuando dice “si q, la recta q, es  $y=2x-3$ , entonces está en -3 [refiriéndose al valor de b]” (E03) y lo señala en el plano cartesiano.

En resumen, la contrastación entre los reportes verbales con los procesos cognitivos anticipados por los especialistas y a su vez la comparación con la especificación del reactivo apunta que este ítem no presenta el problema de sesgo sobre una definición imprecisa del constructo.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMD-11 así como los sondeos que se hicieron durante las entrevistas cognitivas, permitieron encontrar una posible fuente de sesgo en este reactivo.

Tres de los alumnos entrevistados (dos NSE bajo y uno de NSE alto) tuvieron dificultades con el significado de la expresión  $f(x)=mx+b$ , que resulta fundamental para responder de manera correcta el ítem.

Cuando uno de los alumnos de NSE bajo lee el pie del ítem, expresa que le confunde el término  $f(x)=mx+b$ , y que no conoce “esa fórmula”; cuando leyó el reactivo lo hizo de la siguiente manera: “f por X igual a m X más b” (E01).

El otro alumno de NSE bajo, expresó su desconocimiento del término  $f(x)$ , y cuando lo lee lo verbaliza sólo como “equis”, o bien, lo omite en la lectura. Cuando intentó resolver el problema, al parecer no comprendía que  $f(x)$  para este

problema en particular representaba el valor de las ordenadas “y”, por eso cuando hacía una sustitución de la ecuación  $f(x)=2x-3$  para encontrar los puntos en el plano cartesiano, si bien el valor de “x” lo relacionaba con las abscisas, el valor de las ordenadas lo consideraba el valor del “m”, que es 2 (ver Tabla 38).

***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Uno de los entrevistados (docente de NSE medio) aportó elementos en ese sentido y que se describe a continuación (ver Tabla 38).

La docente señala que se requiere enfatizar la recta q, de tal manera que se diferencia del resto de las rectas, para que sea más fácilmente de localizar: “no que todas vayan del mismo grosor, sino que la línea que se está tomando en cuenta sea un poco más oscura que las demás” (E06).

Además, la docente señala que el reactivo tiene otro problema, puesto que si bien la forma general de las rectas es  $f(x)=mx+b$ , la función de la recta q tendría que ser  $y=2X-3$  y no  $f(x)=2x-3$ :

“que aquí [apuntando la parte del pie del reactivo que dice  $f(x)=2x-3$ ] le cambiáramos  $f(x)$  por “y”, porque “y” es la imagen y “y” siempre está en función de “x”.... aquí [apuntando la parte del pie del reactivo que dice  $f(x)=mx+b$ ] se pone f de x porque estamos en función de lo que valga “x”... pero cuando nosotros hablamos ya de  $2x-3$ , ya manejamos “y” para que [sus alumnos ] vean la variable “x” y la variable “y”, o sea, las abscisas y las ordenadas...” (E06).

**Tabla 38. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMD-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 4)	NSE	Bajo	0	2	0	0
			Medio	0	1	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>0</b>
	Sí (N = 5)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	1
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
<b>Total</b>				<b>0</b>	<b>3</b>	<b>0</b>	<b>1</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

### Anexo 30. Análisis de sesgo del reactivo SME-02

El ítem SME-02 (ver Figura 52) pertenece al eje de la asignatura de Matemáticas llamado Sentido numérico y pensamiento algebraico; en el programa de Matemáticas corresponde al tema Significado y uso de las operaciones y al subtema Potenciación y radicación. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de interpretar el significado de elevar un número natural a una potencia de exponente negativo.

Un miligramo se puede expresar como:

$$1 \text{ mg} = 0.001 \text{ g} = 10^{-3} \text{ g}$$

¿Cuál es otra forma equivalente de expresarlo?

(-10)(-10)(-10) g

$\frac{1}{10^3} \text{ g}^*$

(10)(-3) g

$\frac{1}{10(-3)} \text{ g}$

\* Opción correcta

Figura 52. Reactivo SME-02

En la primera etapa de este trabajo de investigación, el SME-02 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.99 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.99 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 13.

De los seis estudiantes que respondieron el reactivo, cuatro de ellos lo respondieron de forma correcta (los dos de NSE bajo y NSE alto) y los dos

alumnos de NSE Medio seleccionaron una opción incorrecta. Por su parte, los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los alumnos que respondieron de forma correcta utilizaron un procedimiento adecuado (ver Tabla 39), en el que: 1) reconocieron que un miligramo es la milésima parte del gramo; 2) identificaron en el pie del ítem las diversas maneras de expresión; y a partir de ahí hubo una bifurcación en el procedimiento, mientras que un alumno utilizó 3a) su conocimiento de las leyes de los exponentes y sabía que en una potencia con número negativo su equivalente es su inverso con potencia positiva, de tal manera que supo que el equivalente de  $10^{-3} \text{ g} = 1/10^3 \text{ g}$ ; en cambio los otros tres alumnos que respondieron correctamente el ítem el siguiente paso en el procedimiento consistió en 3b) desarrollar cada una de las cuatro opciones de respuesta y al hacerlo con la expresión  $1/10^3$  la desarrollaron de la siguiente manera:  $\frac{1}{(10)(10)(10)} = \frac{1}{1000} = 0.001$  y la compararon con una de las equivalencias del miligramo que se presentó en el pie del ítem.

Por su parte, los dos alumnos de NSE medio que seleccionaron un opción incorrecta, utilizaron un procedimiento incorrecto en la resolución del problema, por ejemplo, uno de ellos lo que hizo fue ir comparando cada una de las expresiones equivalentes al miligramo que se presentan en el pie del reactivo con las opciones de respuesta, para descartarlas como correctas. Así, compara 1 mg como equivalente con la respuesta de la opción a) (-10) (-10) (-10) g; luego compara 0.001 g con la opción c) (10)(-3) g; luego compara  $10^{-3}$  g con la opción b)  $1/10^3$  g; y al final selecciona la opción d.

Por su parte, los tres docentes utilizaron un procedimiento de solución apropiado, en el que emplearon su conocimiento sobre las leyes de los exponentes para saber que en una expresión con potencia negativa, su equivalente es el inverso con potencia positiva, que para el ítem es  $10^{-3} \text{ g} = 1/10^3 \text{ g}$ .

**Tabla 39. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SME-02, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 2)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	2	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 4)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>
<b>Total</b>			<b>2</b>	<b>0</b>	<b>0</b>	<b>4</b>	
<b>Fuente:</b> elaboración propia							

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se organizaron según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

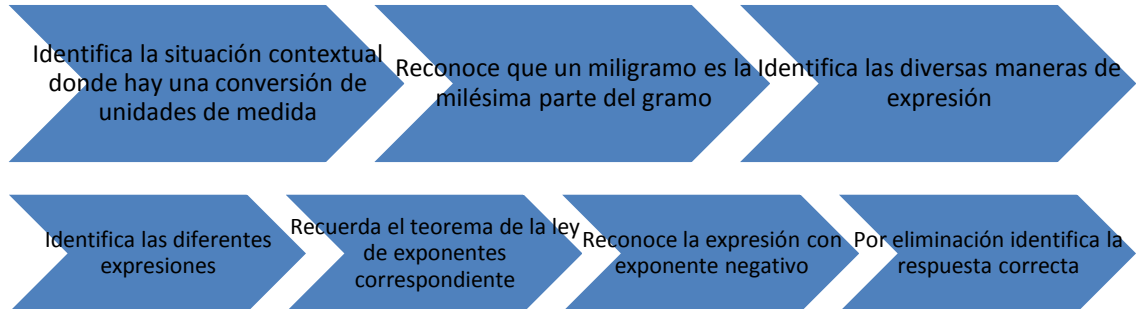
***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

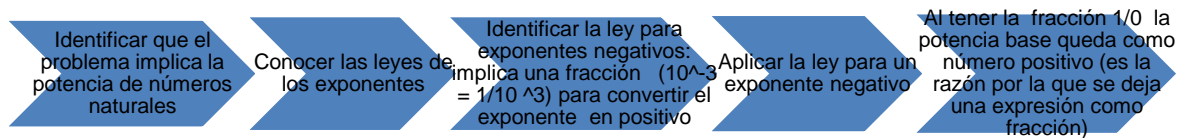
- Multiplicar y dividir números naturales.
- Aplicar las reglas de los signos.
- Aplicar las leyes de los exponentes.
- Potenciar números naturales.
- Resolver ejercicios numéricos.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los senderos cognitivos que se ilustran en la Figura 53:

Proceso 1:



Proceso 2:



**Figura 53. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-02**

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que este ítem tiene coherencia entre los procesos cognitivos que evocarían los sustentantes con el constructo que pretende medir; es decir, cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes podrían utilizar dos senderos cognitivos, sin embargo, en cualquiera de los dos deberían conocer las leyes de los exponentes y emplearlas para identificar una expresión equivalente cuando se tiene un exponente negativo, proceso que coincide con lo que pretende medir el ítem, si el sustentante es capaz de aplicar las leyes de los exponentes y aplicar las reglas de los signos.

Un segundo análisis, a partir de los reportes verbales, permitió conocer que si bien la mayoría de los entrevistados que respondieron correctamente el ítem evocan los conocimientos y habilidades previstos en la especificación del reactivo, tres de los alumnos (dos de NSE bajo y uno de NSE medio), al resolver el reactivo evocan



conocimientos distintos a los señalados en la especificación y emplean operaciones cognitivas diferentes a las anticipadas por las especialistas; es decir, la evidencia apunta a una definición imprecisa del constructo que desea evaluar (ver Tabla 40).

En las entrevistas se identificaron varios segmentos en los que esos tres entrevistados, para resolver correctamente el ítem emplean únicamente sus conocimientos de la división y la multiplicación, y no hay evidencia de que hayan interpretado el significado de elevar un número natural a una potencia de exponente negativo para responderlo.

Por ejemplo, cuando a uno de los alumnos de NSE alto que había seleccionado la opción correcta, se le sondea por qué eligió esa opción, responde “porque... tenemos una división, que si la podemos cambiar sería 1 entre 1000 [mientras escribe 1/1000]”, luego convierte la fracción a notación decimal y verbaliza “...para que esto quepa sería 0.001 [cuando resuelve la división de 1000 entre 1]” (E08); a partir de ese resultado lo compara con el segundo término de las expresiones equivalentes del miligramo que se incluyeron en el pie del ítem: 0.001 g.

Otro alumno, para resolver correctamente el ítem no requirió interpretar el significado de elevar un número natural a una potencia de exponente negativo, más bien, lo hizo a partir de que comprende que tiene que buscar una expresión que sea equivalente a un miligramo (0.001 g), expresión que está planteada en el pie del ítem; a partir sólo de esa expresión y del reconocimiento de algunas de sus características (incluye decimales, es menor a la unidad y tiene signo positivo) descarta las opciones incorrectas. Afirma que la primera opción no es correcta, porque como los tres términos a multiplicar tienen signo negativo, la respuesta sería con signo negativo; y la expresión sobre la que se busca una equivalencia es con término positivo. Descarta la tercera opción señalando que el resultado de la multiplicación sería un 3 con algunos ceros, y la expresión equivalente (o de referencia) es 1. Selecciona la segunda opción (que es la correcta) porque advierte que como producto obtendrá un resultado con decimales y números positivos. Señala que no selecciona la opción d) porque aunque también obtendría como resultado un número con decimales, sería con signo negativo; y eso difiere de la expresión sobre la cual se busca una equivalencia.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SME-02 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE (ver Tabla 40); en resumen, no se encontraron elementos que hagan pensar

que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes.

***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem (ver Tabla 40). Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su desconocimiento de las leyes de los exponentes, o bien, que no emplearon el método de desarrollar cada una de las expresiones incluidas en las opciones de respuesta, para luego comparar su resultado con alguna de las equivalencias del miligramo señaladas en el pie del ítem.

Durante las entrevistas, ninguno de los entrevistados mostró evidencias de que hubo una comprensión inadecuada del pie del ítem o que tuviera problemas de legibilidad. Incluso los alumnos que respondieron de manera incorrecta el ítem, verbalizan una comprensión apropiada de la tarea evaluativa, que se ilustra en la siguiente respuesta que da uno de ellos y cuando la entrevistadora le pregunta si comprendió el reactivo:

“sí, no tenía tanto problema... iba como directo al punto... cuál es la forma equivalente a como está expresado... saca la otra forma en que se puede representar...” (E04).

**Tabla 40. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SME-02, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 2)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 7)	NSE	Bajo	2	0	0	0
			Medio	0	0	0	0
			Alto	1	0	0	0
			<b>Subtotal</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Total</b>				<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

### Anexo 31. Análisis de sesgo del reactivo SME-09

El ítem SME-09 (ver Figura 54) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Diagramas y tablas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas de conteo que se puedan resolver por medio de una multiplicación.

<p>¿De cuántas formas se pueden obtener números de cuatro cifras con los dígitos 1, 2, 3, 4, 5, 6, 7, 8, 9 y 0? Considera que los dígitos se pueden repetir.</p> <p><input type="radio"/> <math>10 \times 9 \times 8 \times 7</math></p> <p><input type="radio"/> <math>10 \times 8 \times 6 \times 4</math></p> <p><input type="radio"/> <math>10 \times 9 \times \dots \times 1</math></p> <p><input type="radio"/> <math>9 \times 10 \times 10 \times 10</math> *</p> <p>* Opción correcta</p>
---

Figura 54. Reactivo SME-09

En la primera etapa de este trabajo de investigación, el SME-09 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo en 0.74 lógitos con respecto a los de NSE alto; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.74 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo, en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 14.

De los seis estudiantes que respondieron el reactivo, ninguno de ellos lo respondió de forma correcta, es decir, los seis alumnos seleccionaron una opción incorrecta o dejaron sin responderlo. Por su parte, de los tres docentes sólo uno (de NSE bajo) lo respondió de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, el docente que lo respondieron de forma correcta utilizaron un

procedimiento adecuado en el que: 1) reconoció que tenía que formar números sólo de cuatro cifras a partir de los dígitos 1, 2, 3, 4, 5, 6, 7, 8, 9 y 0; 2) reconoció que es importante el orden en que forman los números, ya que no es el mismo número 1234 que 4321, es decir, reconoció que se trata de una permutación; 3) reconoció que los dígitos se pueden repetir; 4) estableció que en la primera posición del número sólo pueden emplearse los dígitos del 1 al 9, porque de incluir el cero ya no sería un número de cuatro cifras, sino de tres; 5) estableció que en las posiciones 2, 3 y 4 del número se pueden emplear los 10 dígitos en cada posición; 6) estableció la operación para calcular la cantidad de números que se pueden formar:  $9 \times 10 \times 10 \times 10$

Por su parte, los seis alumnos (de los tres NSE) que seleccionaron una opción incorrecta, desconocían el procedimiento que debían emplear para resolver de forma correcta el ítem, o bien, utilizaron un procedimiento incorrecto en la resolución del problema (ver Tabla 41), por ejemplo, uno de ellos utiliza un procedimiento en el que realiza la multiplicación de cada una de las cuatro opciones de respuesta, anticipando que la que le dé como resultado un número de 4 cifras es la que será la correcta; después realizar las multiplicaciones y obtener en tres de ellas resultados con números de cuatro cifras, se percata que ese procedimiento es incorrecto y deja sin responder el ítem.

Los docentes que no seleccionaron la opción correcta, en cambio, a pesar de que conocían cómo realizar permutaciones, no consideraron un segmento del procedimiento, en el que tenían que distinguir que en la primera posición de los posibles números de cuatro cifras sólo pueden emplearse los dígitos del 1 al 9, porque de incluir el cero ya no sería un número de cuatro cifras, sino de tres. Esta situación se abordará más adelante como un posible problema del ítem.

**Tabla 41. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SME-09, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 5)	Nivel Socioeconómico	Bajo	2	0	0	--
			Medio	2	0	0	--
			Alto	2	0	0	--
			<b>Subtotal</b>	<b>6</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 1)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	0	0	0
			Alto	--	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	<b>Total</b>			<b>6</b>	<b>0</b>	<b>0</b>	<b>0</b>
	<b>Fuente:</b> elaboración propia						

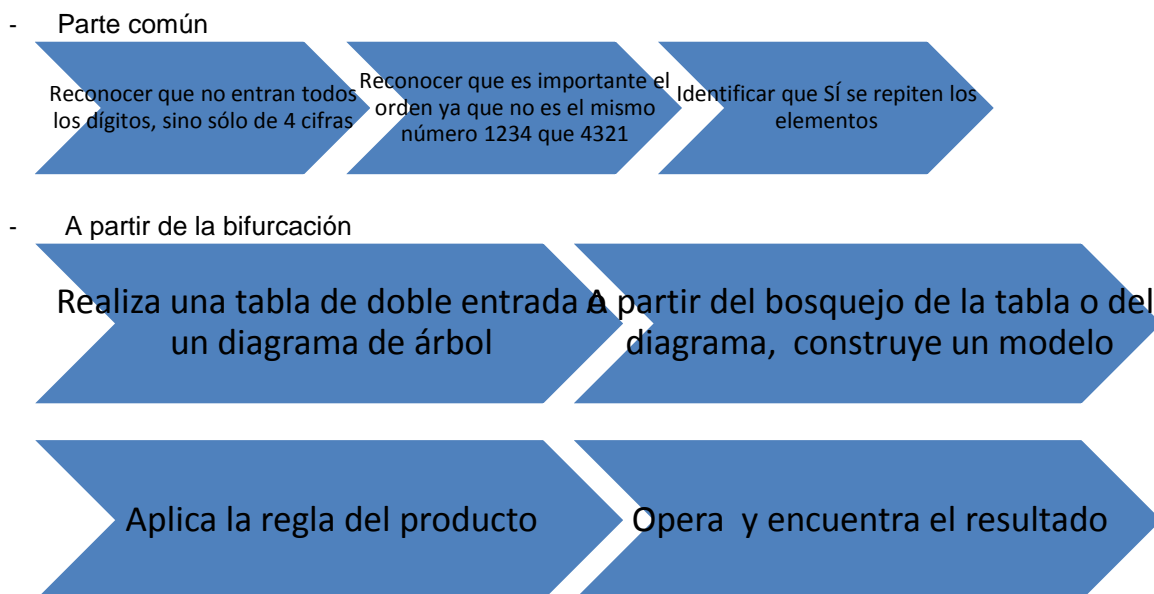
En los subapartados siguientes se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente, los cuales están organizados por cada uno de los tres Elementos del Diseño Universal de las Evaluaciones que se utilizaron como categorías de análisis: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Problemas de conteo y sus diferentes representaciones.
- Uso de diagramas de árbol y tablas de doble entrada.
- Conocer y comprender la regla del producto.
- Analizar las diferentes posibilidades planteadas.
- Identificar el producto que resuelve el problema.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los dos senderos cognitivos que se ilustran en el siguiente esquema, los cuales tienen una parte común (ver Figura 55):



**Figura 55. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-09**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas, se encontró coherencia entre ambas fuentes, lo que apunta a que los procesos que tiene que evocar para responder correctamente este ítem son coherentes con el constructo que pretende medir; cuando las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían reconocer que el problema implica una permutación con remplazo con números de 4 cifras, eso es coincidente con lo que la especificación señala que el ítem pretende medir: si el sustentante conoce y comprende la regla del producto; sabe analizar las diferentes posibilidades planteadas; e identifica el producto que resuelve el problema.

Además, un análisis más detallado, a partir de los reportes verbales, corroboró que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean las operaciones cognitivas anticipadas por las especialistas (ver Tabla 42).

En otras palabras, la evidencia disponible muestra que cuando los entrevistados responden de forma correcta el reactivo tienen que: identificar que el problema es de conteo; reconocer que tienen que formar números de cuatro cifras con los dígitos 1, 2, 3, 4, 5, 6, 7, 8, 9 y 0; reconocer que se trata de una permutación donde los dígitos se pueden repetir; y establecer la cantidad de dígitos que se

pueden colocar en cada una de las cuatro posiciones de los posibles números de cuatro cifras. Es decir, no se encontró evidencia empírica que indicase que este ítem presenta un problema de una definición imprecisa del constructo, por el contrario, la verbalización del alumno que seleccionó la opción correcta, ilustra la evocación de los conocimientos y habilidades señalados, cuando dice:

“entonces aquí para ver cuántos números de cuatro cifras... sería una manera de multiplicarlos entre sí... y luego la cuarta, supongo, porque son cuatro cifras... []... sí, porque si fueran un número de 2 cifras sería  $10 \times 10$ ...” (E07).

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SME-09 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE (ver Tabla 42); en resumen, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes. De hecho, el único alumno que lo resolvió es de NSE alto, cuando a partir de los resultados del análisis de la primera etapa de la investigación este ítem se diagnosticó con un DIF favorable a los alumnos de NSE bajo.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Cuatro de los entrevistados (los tres docentes y un alumno de NSE bajo) aportaron elementos en ese sentido (ver Tabla 42).

El docente de NSE bajo, quien respondió correctamente el ítem, al ir verbalizando el proceso cognitivo que empleaba, dice:

“Veo un problema en ésta [pregunta]... no les dice que el primer dígito no puede ser cero... dice ‘considera números de cuatro dígitos’ y entonces aquí, al elegir el primer número, el primer dígito, son diez opciones, al escribir el segundo tienen otras 10 opciones, el tercero otras 10 opciones y el cuarto otras 10 opciones... y ninguna de las respuestas probables representa esto...” (E03).

Luego la entrevistadora lo sondea con “¿y entonces cómo la razonó usted?”, a lo que el docente responde “el primer dígito no puede ser cero, entonces nada más tengo nueve opciones para el primero” (E03). La entrevistadora plantea una nueva pregunta “¿y considera que esa sería una dificultad?”, a lo que el profesor



responde “esa sería una dificultad para el alumno, creo que los niños no lo relacionan de esa manera, entonces tendría que decir que el primer dígito no puede ser cero” (E03). Además, el docente argumenta que se podría escribir “0123” y sería un número de 4 cifras, que se sabe que el cero no tomaría valor, pero que sigue siendo un número de cuatro dígitos que es lo que plantea el problema.

Los otros dos docentes, se sorprendieron cuando bajo su razonamiento la respuesta correcta debería ser 10X10X10X10 y no la encontraron dentro de las opciones. Es decir, si bien los docentes sabían que el problema implicaba una permutación donde los dígitos se puede repetir, el reactivo tiene un elemento de complejidad innecesaria, más asociado a la Lógica, en la que los sustentantes deben considerar que en el conjunto de los números de cuatro cifras, aquellos que inicien con 0 no suele considerarse como parte de ese conjunto.

**Tabla 42. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SME-09, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 8)	NSE	Bajo	0	0	0	1
			Medio	0	0	0	1
			Alto	0	0	0	1
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>3</b>
	Sí (N = 1)	NSE	Bajo	0	0	0	1
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
	<b>Total</b>			<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>

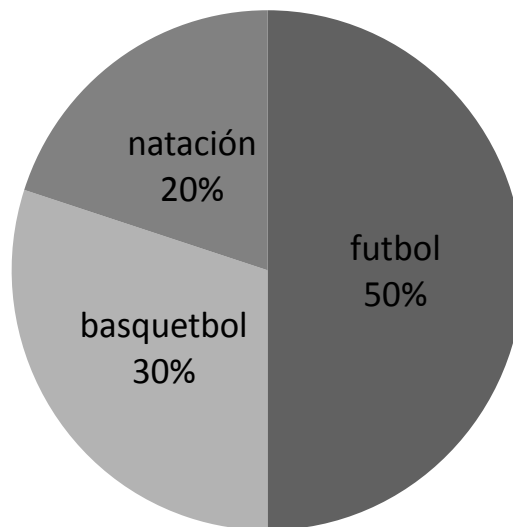
**Notas:** NSE=Nivel Socioeconómico

**Fuente:** elaboración propia

### Anexo 32. Análisis de sesgo del reactivo SME-11

El ítem SME-11 (ver Figura 56) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Gráficas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen la interpretación de información representada en gráficas de barras o circulares de frecuencia absoluta o relativa expresada como fracción, decimal o porcentaje.

La gráfica muestra la distribución de 120 estudiantes de tercero de secundaria que practican tres diferentes deportes.



El 30% de los estudiantes que practican fútbol son mujeres. ¿Cuántas mujeres practican fútbol?

- 60
- 30
- 18 \*
- 15

\* Opción correcta

Figura 56. Reactivo SME-11

En la primera etapa de este trabajo de investigación, el SME-11 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.75 lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.75 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 15.

De los seis estudiantes que respondieron el reactivo, la mitad de ellos lo respondió de forma correcta (los dos estudiantes de NSE alto y uno de NSE medio) y los tres alumnos restantes (dos de NSE bajo y una de NSE medio) seleccionaron una opción incorrecta o dejaron sin responderlo. Por su parte, los tres docentes lo respondieron de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los alumnos que lo respondieron de forma correcta utilizaron un procedimiento adecuado (ver Tabla 43), en el que: 1) identificaron que la gráfica circular expresa información en porcentajes; 2) reconocieron que los porcentajes de la gráfica que expresan la proporción de estudiantes que practica cada uno de los tres deportes está en función de los 120 de estudiantes de tercero de secundaria; 3) reconocieron que para resolver el problema sólo es necesaria la información correspondiente al fútbol y para ello algunos calcularon mentalmente que la mitad de los 120 estudiantes equivale a 60 jugadores de fútbol, mientras que otros hicieron una regla de tres y llegaron al mismo resultado; 4) calcularon la cantidad de mujeres que juegan fútbol tomando en cuenta que ahora el 100% equivale a los 60 jugadores de fútbol, y lo hicieron a partir de una regla de tres, obteniendo como resultado una cantidad de 18 jugadoras de fútbol.

Por su parte, de los tres alumnos que seleccionaron una opción incorrecta, una de ellas (de NSE medio) utilizó un procedimiento incorrecto en la resolución del problema, pues realizó una serie de multiplicaciones y divisiones bajo un razonamiento un tanto azaroso, en el que esperaba obtener como resultado alguna de las opciones de respuesta; por otro lado, los dos alumnos de NSE bajo, si bien conocían que debían emplear una regla de tres para enfrentarse a la tarea evaluativa, no la operaron de forma correcta, o bien, no consideraron que el problema planteaba un doble cálculo de proporcionalidad, y sólo hicieron uno de ellos, por lo cual no consiguieron un resultado correcto (ver Tabla 43).

Los docentes también emplearon un procedimiento de resolución adecuado, similar al que utilizaron los alumnos que respondieron de forma correcta el ítem.

**Tabla 43. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SME-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 3)	Nivel Socioeconómico	Bajo	0	0	2	--
			Medio	1	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>0</b>
	Sí (N = 3)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	0	0	1
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>3</b>
<b>Total</b>			<b>1</b>	<b>0</b>	<b>2</b>	<b>3</b>	

Fuente: elaboración propia

Enseguida se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según el problema potencial que se indagó, que coincide con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

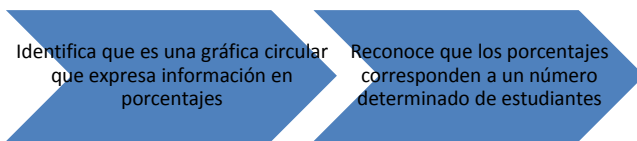
### ***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

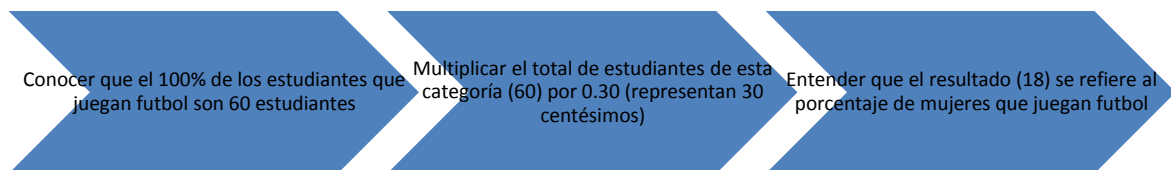
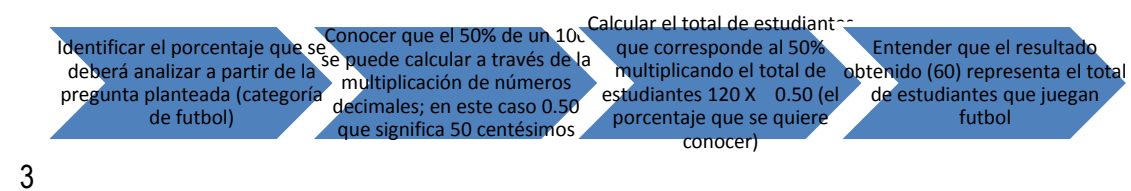
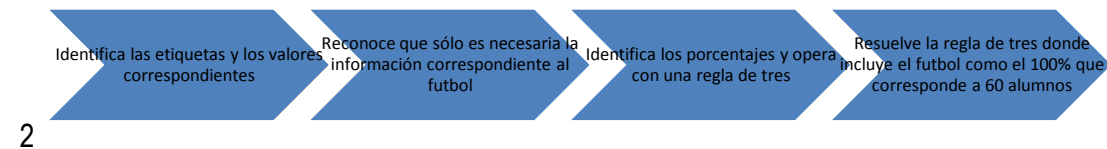
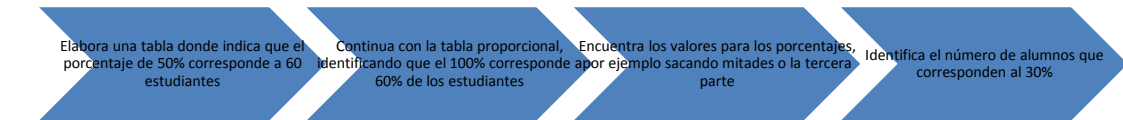
- En gráficas circulares: cálculo de porcentajes e identificar categorías e interpretación de los datos.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los tres senderos cognitivos que se ilustran la Figura 57, los cuales tienen una parte común:

- Parte común



- A partir de la trifurcación



**Figura 57. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-11**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas se encontró consistencia entre ambas fuentes. Cuando las especialistas anticiparon que para resolver correctamente el ítem, los sustentantes podrían ser alguno de los cuatro senderos cognitivos de la figura anterior, un proceso común a todos y que resulta fundamental para enfrentarse al ítem es el cálculo de porcentajes dentro de una de las categorías de deportes (futbol) y eso es coincidente con lo que el ítem pretende medir: si el sustentante, en gráficas circulares, es capaz de calcular porcentajes e identificar categorías e interpretación de los datos.

Un segundo análisis más detallado, a partir de los reportes verbales, corroboro que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean algunos de los senderos cognitivos anticipados por las especialistas (ver Tabla 44). En otras palabras, la evidencia disponible muestra que cuando los entrevistados responden de forma correcta el reactivo tienen que: identificar categorías e interpretar los datos en gráficas circulares así como calcular porcentajes e incluso porcentajes relativos para aquellos que realizaron dos reglas de tres para encontrar la respuesta correcta. Y eso significa que este reactivo no presenta el problema de definición imprecisa del constructo.

La verbalización de uno de los alumnos que seleccionó la opción correcta, ilustra la evocación de los conocimientos y habilidades señalados, cuando dice:

“si 60 personas juegan futbol, y me dice que el 30% de los estudiantes que juegan futbol son mujeres, 60 es 100 por ciento [al tiempo que escribe en el cuadernillo:  $\frac{60}{-30\%} = 100\%$ ].... Estoy buscando cuántas personas son el 30%, entonces lo haría por regla de tres” (E08).

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SME-11 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE (ver Tabla 44). Si se considera que durante la primera etapa de esta investigación este reactivo tuvo un DIF severo que favorece a los alumnos de NSE alto, y si bien en la segunda etapa de este trabajo los alumnos de NSE alto son los que lo respondieron de forma correcta y con los entrevistados de NSE bajo no ocurrió así, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes, sino que para este reactivo lo que hizo la diferencia fue la comprensión a cabalidad de los conocimientos y habilidades establecidos en la especificación del reactivo.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas aporta elementos para afirmar sobre dos posibles problemas de incomprensibilidad o legibilidad del ítem (ver Tabla 44). Dos de los docentes (que trabajan con alumnos de NSE bajo y NSE medio) aportaron elementos en ese sentido (ver Tabla 44).

El docente que atiende alumnos de NSE bajo, advirtió sobre un posible problema que no haya uniformidad en la manera en que se escriben las cantidades en el pie del reactivo, y eso podría distraer a los alumnos:

“pueden parecer detalles menores, pero a veces... es que los números los escriben con número y luego de pronto se los escriben con letra, aquí el ‘120’ y luego aquí ‘tres’ con letra, y de pronto eso lo hace distraerse un poco...” (E03).

Vale decir que este posible problema sobre una eventual distracción de los alumnos, no se percibió con ninguno de los alumnos entrevistados.

La docente que atiende alumnos de NSE medio, advirtió sobre un segundo problema potencial: que se utilice el mismo valor de un porcentaje que corresponde a información distinta en el pie del reactivo; en la gráfica circular el deporte de basquetbol lo practica el 30% de los estudiantes y el problema plantea calcular a cuánto equivale el 30% de las mujeres que juegan futbol, lo cual podría confundir a los alumnos, o bien, debido a una distracción, les podría resultar confuso discernir ambos datos:

“...aquí se puede confundir el niño, porque va a decir ‘está mal la respuesta’ porque puede decir ‘el 30% no, practican el basquetbol’... entonces ese el chiste que el niño sepa distinguir que son del futbol y no del básquetbol” (E06).

La docente recomienda que se cambien los valores de los porcentajes de natación y basquetbol, para que ninguno de los dos sea 30.

**Tabla 44. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SME-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 3)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 6)	NSE	Bajo	0	0	0	1
			Medio	0	0	0	1
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>
	<b>Total</b>			<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia



### **Anexo 33. Análisis de sesgo del reactivo SME-12**

El ítem SME-12 (ver Figura 58) pertenece al eje de la asignatura de Matemáticas llamado Manejo de la información; en el programa de Matemáticas corresponde al tema Representación de la información y al subtema Gráficas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas que impliquen la interpretación de información representada en gráficas de línea.

En la primera etapa de este trabajo de investigación, el SME-12 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.79 lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.79 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

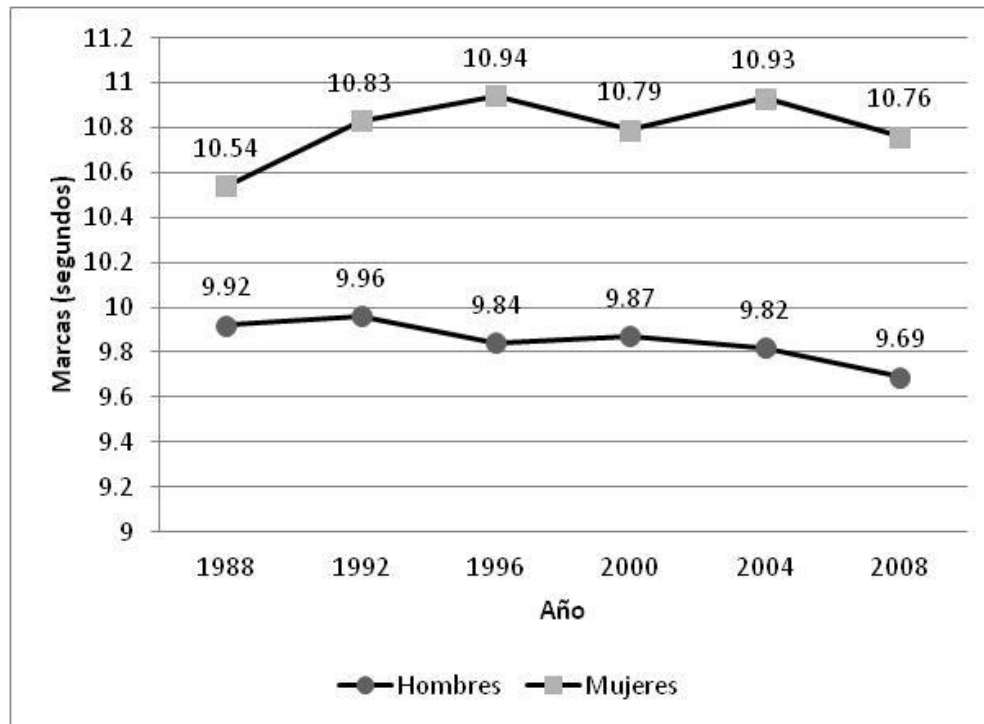
El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 16.

De los seis estudiantes que respondieron el reactivo, sólo uno de ellos no lo respondió de forma correcta (de NSE medio). Por su parte, los tres docentes lo respondieron de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, de los cinco alumnos que lo respondieron de forma correcta cuatro de ellos (de NSE alto y NSE bajo) utilizaron un procedimiento adecuado (ver Tabla 45), en el que: 1) reconocieron cada una de las diferentes etiquetas de la gráfica de líneas: marcas olímpicas en segundos, año de los juegos olímpicos y todo ello desagregado para hombres y mujeres; 2) comprendieron que el problema consiste en tener que encontrar el menor tiempo de las atletas mujeres y restárselo al menor tiempo de los hombres; 3) identificaron en la gráfica que el menor tiempo logrado por las mujeres fue de 10.54 segundos en 1988; 4) identificaron que el menor tiempo logrado por los hombres fue de 9.69 segundos en los juegos olímpicos de 2008; 5) operaron la resta  $10.54 - 9.69 = 0.85$  segundos.

El alumno de NSE medio que seleccionó la respuesta correcta, aparentemente adivinó, pues su proceso de resolución consistió en probar con múltiples restas en las que incluyo diversos tiempos de hombres y mujeres, hasta que encontrara una donde consiguiera alguno de los resultados de las opciones de respuesta.

La siguiente gráfica muestra las marcas olímpicas alcanzadas en los 100 m planos por atletas del sexo masculino y femenino en los últimos seis Juegos Olímpicos.



Identifica el menor tiempo logrado por las mujeres y compáralo con el menor tiempo logrado por los hombres. ¿Cuál es la diferencia entre estas marcas olímpicas?

- 10.54 s
- 9.69 s
- 8.5 s
- 0.85 s \*

\* Opción correcta

Figura 58. Reactivo SME-12

Por su parte, la alumna (de NSE medio) que seleccionó una opción incorrecta, si bien aportó evidencias de que comprendía que debía calcular la diferencia entre los tiempos de las mujeres y los hombres, hubo algunos elementos del proceso de solución que desconocía, por ejemplo, no supo identificar en la gráfica cuál era el

menor tiempo conseguido tanto por los atletas del sexo femenino como masculino (ver Tabla 45).

Los docentes también emplearon un procedimiento de resolución adecuado, similar al que utilizaron los alumnos que respondieron de forma correcta el ítem.

**Tabla 45. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SME-12, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 1)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	0	0	1	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>
	Sí (N = 5)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	1	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>4</b>
<b>Total</b>				<b>0</b>	<b>1</b>	<b>1</b>	<b>4</b>
<b>Fuente:</b> elaboración propia							

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se organizaron por cada uno de los tres Elementos del Diseño Universal de las Evaluaciones analizados en este trabajo: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

### ***Definición imprecisa del constructo***

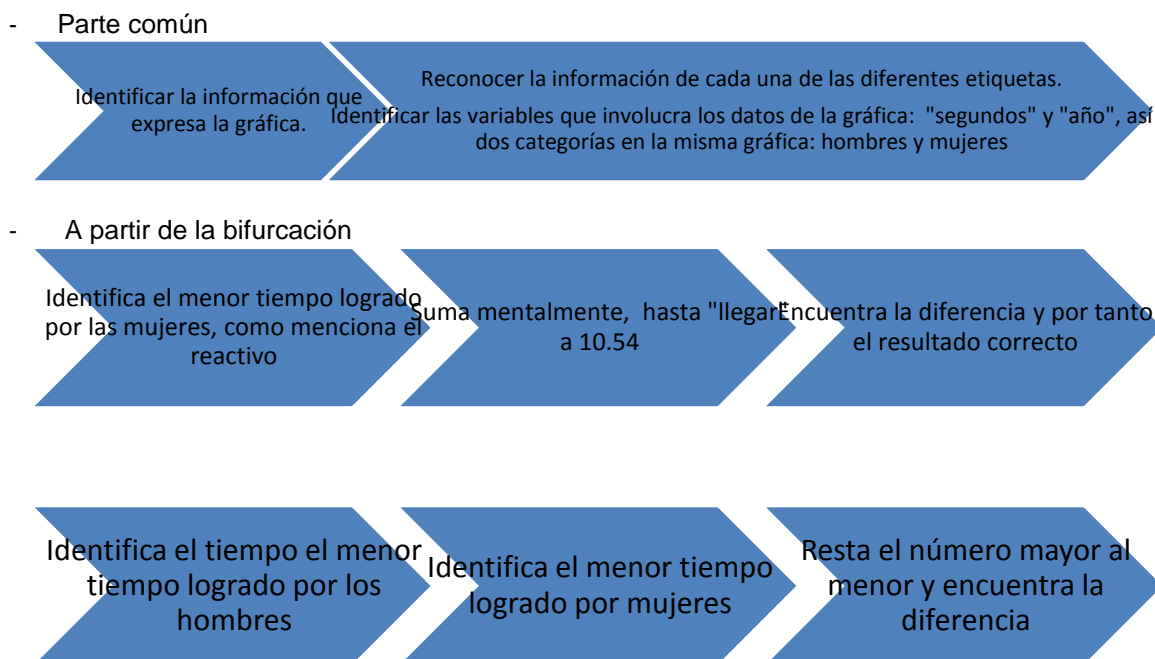
Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Comprender la información contenida en la gráfica. La comprensión de una gráfica es un proceso complejo que implica identificar claramente los elementos estructurales que, según Friel, Curcio y Bright (2001), son:

- Contenido contextual: Saber leer el título y etiquetas (para comprender qué información se muestra en la gráfica y cuáles son las variables representadas).
- Marco del gráfico: Saber identificar los ejes e interpretar correctamente la escala usada en cada uno (para comprender las unidades de medida de las magnitudes representadas en el gráfico). Por lo que la interpretación de números (naturales, fracciones comunes decimales) en la recta numérica es un conocimiento esencial para dar respuesta al reactivo.
- Especificadores del gráfico: Saber interpretar que, en el caso de un gráfica de línea, las coordenadas de los puntos unidos por los segmentos que forman la línea, son parejas ordenadas cuyo primer elemento corresponde a la unidad de tiempo y el segundo elemento corresponde a la magnitud graficada y representada en el eje vertical. Esto implica que los alumnos estén familiarizados con la representación de puntos en un plano cartesiano.
- Para el caso del contenido que se evaluará y como se presentará una gráfica de línea en con dos magnitudes que varían con respecto al tiempo, para responder al reactivo se tendrán que reconocer los elementos estructurales mencionados en ambas magnitudes por separado y, además, relacionándolos entre sí.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los dos senderos cognitivos que se ilustran en la Figura 59, los cuales tienen una parte común.

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó a que existe coherencia entre ambas fuentes; mientras que la especificación del reactivo indica que lo que se pretende evaluar es si el sustentante es capaz de leer e interpretar la información contenida en una gráfica que incluye varias variables, así como las escalas de los ejes del plano cartesiano, las especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían interpretar de manera adecuada la gráfica, con los mismos elementos señalados en la especificación pero acotados al problema en particular que el ítem requería resolver.



**Figura 59. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SME-12**

Un segundo análisis en profundidad, a partir la información derivada de las entrevistas cognitivas, sugiere que si bien la mayoría de los entrevistados (estudiantes y docentes) al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean algunos de los senderos cognitivos anticipados por las especialistas, hubo un par de alumnos (de NSE bajo y NSE medio) que aportaron evidencias de una posible imprecisión del constructo (ver Tabla 46).

Esos dos alumnos, si bien respondieron de forma correcta del ítem y evocaron algunos de los conocimientos que plantea la especificación del reactivo, no requirieron evocar otros que resultan fundamentales y aún así dedujeron la respuesta correcta.

Por ejemplo, una alumna resolvió el ítem a partir de la Lógica; ella fue capaz de distinguir el menor tiempo de las mujeres en los juegos olímpicos de 1988; sin embargo, no fue capaz de identificar el menor tiempo de los hombres que ocurrió en los juegos olímpicos de 2008, sino que utilizó el tiempo de los hombres del mismo año que el de las mujeres (1988). Después de realizar la resta  $10.54 - 9.92 = 0.62$  y comparar su resultado con las opciones de respuesta y no coincidir con ninguna, selecciona la opción d), que es correcta. Cuando se le pide que justifique porqué escogió esa opción, responde:

“porque no puede haber tanta diferencia... el de las mujeres es 10.54 y el de los hombres es 9.92, [otra opción] sería mucha diferencia” (E02).

Al similar ocurrió con otro alumno, que si bien evoca algunos de los conocimientos que se pretende medir (como identificar las marcas de los hombres y las mujeres, así como distinguir que la gráfica presenta información para distintos años), es evidente que el alumno está adivinando porque realizó cuatro restas antes de seleccionar la opción correcta, en cada resta utilizaba como minuendo alguna de las marcas de las mujeres y en el sustraendo alguna de los atletas masculinos; cuando la entrevistadora le pide que le explique lo que está realizando con las restas, el alumno responde “estoy viendo si puedo encontrarme la respuesta... por ejemplo escogí esta opción del más bajo de los hombres se lo voy a sumar [al dato de la opción de respuesta d)] para ver cuánto me da, si llega a acercarse cuando menos” (E05), luego realiza la siguiente suma  $9.69 + 0.85 = 10.54$ , y luego señalando el valor en la gráfica dice “se acerca a éste”; a partir de ahí realiza la resta  $10.59 - 9.69 =$  y sin escribir el resultado selecciona la opción d.

Es decir, para un sustentante que ha comprendido que debe realizar una resta de los valores que se presentan en la gráfica, de las cuatro opciones de respuesta la única que es plausible, es la opción correcta de 0.85 segundos; cualquier otra diferencia que se realice con los tiempos de la gráfica no generará un resultado mayor o igual a 8.5 segundos, que es la siguiente opción con mayor tiempo después de la opción correcta.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SME-12 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE (ver Tabla 46). Si se considera que durante la primera etapa de esta investigación este reactivo tuvo un DIF severo que favorece a los alumnos de NSE alto, sin embargo, en la segunda etapa de este trabajo los alumnos tanto de NSE alto como NSE bajo lo respondieron de forma correcta, además, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes, sino que para este reactivo lo que hizo la diferencia fue la comprensión a cabalidad de los conocimientos y habilidades establecidos en la especificación del reactivo.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem (ver Tabla 46). Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su desconocimiento para interpretar de forma adecuada los diferentes elementos que contenía la gráfica, tales como: marcas de tiempo en segundos, años de los juegos olímpicos y esa información desagregada para hombres y mujeres.

Durante las entrevistas, ninguno de los entrevistados mostró evidencias de que hubo una comprensión inadecuada del pie del ítem o que tuviera problemas de legibilidad. Incluso los alumnos que respondieron de manera incorrecta el ítem, verbalizan una comprensión apropiada de la tarea evaluativa, que se ilustra en la siguiente respuesta que da uno de ellos y cuando la entrevistadora le pregunta si comprendió el reactivo:

“...sí sé qué es lo que me está pidiendo [el reactivo]... cuál es la diferencia de un tiempo, cuánto es lo que se llevan...” (E04).

**Tabla 46. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SME-12, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

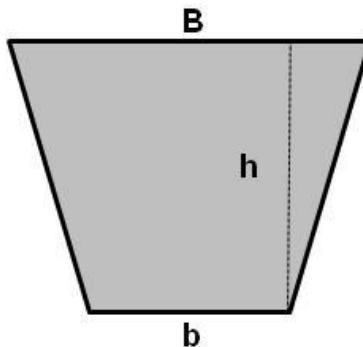
				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 1)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 8)	NSE	Bajo	1	0	0	0
			Medio	1	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Total</b>				<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

### Anexo 34. Análisis de sesgo del reactivo SMF-02

El ítem SMF-02 (ver Figura 60) pertenece al eje de la asignatura de Matemáticas llamado Sentido numérico y pensamiento algebraico; en el programa de Matemáticas corresponde al tema Significado y uso de las literales y al subtema Patrones y fórmulas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de traducir al lenguaje natural el significado de fórmulas geométricas o viceversa.

Observa que las letras **B**, **b** y **h** representan las medidas del trapecio.



Pablo dice que la fórmula para calcular su área es  $A = \frac{(B+b)h}{2}$ . ¿Cómo puede interpretarse esta fórmula?

- El área de un trapecio es igual al cuadrado de su base por la altura entre dos.
- El área de un trapecio es igual a la suma de la base mayor más la base menor; el resultado de esta suma se multiplica por su altura y se divide entre dos.\*
- El área de un trapecio es igual al producto de la base mayor por la base menor por la altura entre dos.
- El área de un trapecio es igual a la suma de la base mayor más la base menor más la altura entre dos.

\* Opción correcta

Figura 60. Reactivo SMF-02

En la primera etapa de este trabajo de investigación, el SME-12 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto en 0.74



lógitos con respecto a los de NSE bajo; es decir, comparando alumnos con un mismo nivel de habilidad, el reactivo resultó 0.74 lógitos más fácil para aquellos cuyo origen socioeconómico es Alto, en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 17.

De los seis estudiantes que respondieron el reactivo, sólo uno de ellos no lo respondió de forma correcta (de NSE medio). Por su parte, los tres docentes lo respondieron de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los cinco alumnos que lo respondieron de forma correcta utilizaron un procedimiento adecuado (ver Tabla 47), en el que: 1) conocen el uso de los signos operatorios (suma, resta, división, multiplicación) y el de la igualdad; 2) identificaron las partes que conforman el trapecio; 3) relacionaron las literales propuestas con las partes del trapecio; y 4) interpretaron el significado de las literales que dan sentido a la fórmula del área de un trapecio

Por su parte, la alumna de NSE medio que seleccionó una opción incorrecta utilizó un procedimiento de solución incorrecto, en el que le asignó valores a la fórmula del trapecio para intentar obtener su área (ver Tabla 47).

**Tabla 47. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMF-02, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 1)	Nivel Socioeconómico	Bajo	0	0	0	--
			Medio	1	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 5)	Nivel Socioeconómico	Bajo	--	0	0	2
			Medio	--	0	0	1
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>5</b>
	<b>Total</b>			<b>1</b>	<b>0</b>	<b>0</b>	<b>5</b>
	<b>Fuente:</b> elaboración propia						

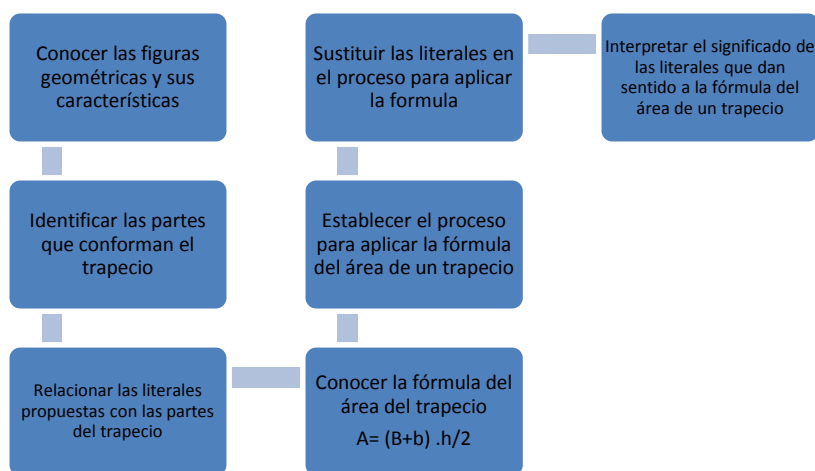
A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente, los cuales se organizaron por cada uno de los tres Elementos del Diseño Universal de las Evaluaciones que se analizaron en esta investigación: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

**Definición imprecisa del constructo**

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Conocimiento de perímetros y áreas de figuras.
- Suma, resta y multiplicación con números naturales, decimales y fraccionarios.
- Conocer el uso de los signos operatorios y el de la igualdad.
- Interpretar una letra como un número general.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un sendero cognitivo como el que se ilustra en la Figura 61:



**Figura 61. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMF-02**

Al contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas se encontró consistencia entre ambas fuentes; es decir, mientras que las

especialistas anticiparon que para enfrentarse a la tarea evaluativa los sustentantes deberían tener conocimiento de figuras geométricas y sus características e identificar las partes que conforman el trapecio, un elemento fundamental es que los sustentantes pudieran relacionar las literales propuestas con las partes del trapecio, entre otras cuestiones para interpretar el significado de las literales que dan sentido a la fórmula del área de un trapecio, lo cual es coincidente con lo que la especificación del ítem señala que busca evaluar, si el sustentante tiene conocimiento de perímetros y áreas de figuras, conoce el uso de los signos operatorios y el de la igualdad, y además, si es capaz de interpretar una letra como un número general.

Un segundo análisis, a partir de la información de las entrevistas cognitivas permitió corroborar que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con ciertas variantes, las operaciones cognitivas anticipadas por las especialistas, donde la operación cognitiva medular consiste en interpretar la fórmula, sus elementos y las operaciones a realizar con ellos.

Por ejemplo, cuando algunos alumnos al ir leyendo el pie del reactivo y llegan a la parte donde tienen que leer la fórmula del trapecio que plantea Pablo  $A = \frac{(B+b)h}{2}$  lo hacen de manera correcta sin siquiera haber leído las opciones de respuesta: “base mayor mas base menor por altura sobre dos” (E01, E03, E05... E09).

Otra cita que ilustra cómo los sustentantes evocan los conocimientos previstos por la especificación, es cuando a un alumno se le pide que explique paso a paso cómo llegó a la respuesta que seleccionó, que es correcta, y explica cómo fue discriminando las opciones:

“primero lo que me estaba diciendo aquí [señalando la opción a] y viendo la fórmula, entonces iba leyendo ‘el trapecio es igual al cuadrado de su base’ y aquí no hay nada que esté indicando que la base sea al cuadrado, entonces esa la descarto; después ésta [señalando la opción c] ‘el área de un trapecio es igual al producto de la base...’ y se suman; aquí iba bien [señalando la opción d], y dice ‘más la altura’ y la altura está multiplicando esta suma” (E08).

O bien, cuando otro alumno al ir leyendo cada una de las opciones de respuesta, llega el turno de leer la segunda opción de respuesta y verbaliza: “esto es lo que yo acabo de decir” (E01), refiriéndose a la manera en que leyó la fórmula del trapecio en el pie del reactivo.

Toda esta evidencia indica que el este ítem no presenta un problema de definición imprecisa del constructo.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMF-02 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE. Si se considera que durante la primera etapa de esta investigación este reactivo tuvo un DIF severo que favorece a los alumnos de NSE alto, sin embargo, en la segunda etapa de este trabajo los alumnos tanto de NSE alto como NSE bajo lo respondieron de forma correcta, además, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes, sino que para este reactivo lo que hizo la diferencia fue la comprensión a cabalidad de los conocimientos y habilidades establecidos en la especificación del reactivo.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando ocurrió el caso de que una alumna seleccionara una opción incorrecta, se debió principalmente a que no fue capaz de interpretar de manera adecuada la fórmula del trapecio.

Cuando a uno de los alumnos se le pide que explique con sus propias palabras lo que el ítem le pide que realice, responde:

“nada más esto [señalando la fórmula del trapecio] cambiarlo a palabras, entonces voy a leer todas [las opciones] para ver cuál es la que representa la fórmula”

Por su parte, cuando se les sondeó a los docentes si habían detectado algún problema sintáctico en el ítem o alguna palabra que causara confusión, los tres señalaron que les parecía adecuada la redacción del reactivo, “no, no tiene ninguna complicación, está bien” (E06).

### Anexo 35. Análisis de sesgo del reactivo SMG-07

El ítem SMG-07 (ver Figura 62) pertenece al eje de la asignatura de Matemáticas llamado Forma, espacio y medida; en el programa de Matemáticas corresponde al tema Medida y al subtema Estimar, medir y calcular. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de identificar el aumento o disminución del volumen al variar la longitud de una arista de un cubo.

Se tienen dos velas en forma de cubo, una rosa y otra verde. La verde tiene una arista de 2 cm y la rosa de 6 cm. ¿Cuántas veces es mayor el volumen de la vela rosa comparado con el de la verde?

- 3 veces
- 9 veces
- 12 veces
- 27 veces\*

\* Opción correcta

**Figura 62. Reactivo SMG-07**

En la primera etapa de este trabajo de investigación, el SMG-07 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico alto con 0.81 lógitos con respecto a los alumnos de NSE bajo; es decir, comparando alumnos con el mismo nivel de habilidad, el reactivo resultó 0.81 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto en comparación con los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 18.

De los seis estudiantes que respondieron el reactivo, sólo dos lo respondieron de forma correcta (de NSE medio y NSE alto), mientras que los cuatro estudiantes restantes (de los tres NSE), lo respondieron de forma errónea. Por su parte, los tres docentes lo respondieron de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los dos alumnos que lo respondieron de forma correcta utilizaron un procedimiento adecuado (ver Tabla 48), en el que: 1) reconocieron que es el problema implicaba la comparación del volumen de dos cubos de diferentes aristas; 2) conocían que para obtener el volumen de un cuerpo geométrico se debe multiplicar sus tres dimensiones: longitud X ancho X altura y que para este caso, al tratarse de un cubo, la arista tiene la misma longitud para esa tres dimensiones; 3) calcularon el volumen de cada vela; y 4) dividieron el volumen de la vela rosa entre el volumen de la vela verde, para conocer cuántas veces cabe la segunda en la primera. De la misma manera, los tres docentes emplearon un procedimiento similar.

Por su parte, los cuatro alumnos que seleccionaron una opción incorrecta utilizaron un procedimiento de solución incorrecto, en el que omitieron calcular los volúmenes de cada vela, y dividieron la mayor de las aristas entre la menor de ellas; o bien, una alumna en lugar de calcular el volumen, calculó el área de cada lado de las velas y luego realizó la división del área mayor entre el área menor, y al no coincidir su resultado con alguna de las opciones de respuesta, empleó el mismo procedimiento que los demás, donde sólo comparó el tamaño de las aristas (ver Tabla 48).

**Tabla 48. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMG-07 desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 4)	Nivel Socioeconómico	Bajo	2	0	0	--
			Medio	1	0	0	--
			Alto	1	0	0	--
			<b>Subtotal</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 2)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	0	0	1
			Alto	--	0	0	1
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>
<b>Total</b>			<b>4</b>	<b>0</b>	<b>0</b>	<b>2</b>	
<b>Fuente:</b> elaboración propia							

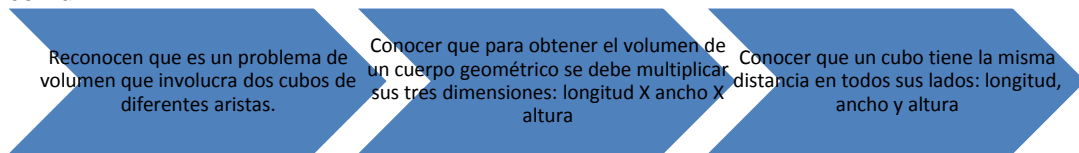
En los siguientes párrafos se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Los resultados se organizan por cada uno de los tres Elementos del Diseño Universal de las Evaluaciones que se utilizaron como categorías de análisis en esta tesis: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

**Definición imprecisa del constructo**

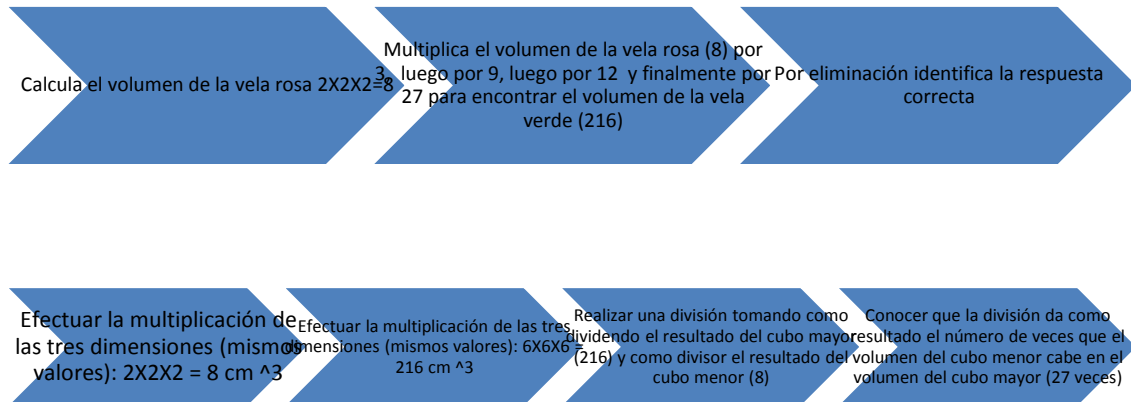
Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante sea capaz de justificar las fórmulas para calcular el volumen de un cubo y de su área lateral o total, así como calcular el área de cuadrados, lo que implica a su vez, que sea capaz de obtener potencias cuadradas de un número.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un sendero cognitivo que tiene una parte común y luego podría tomar alguno de los dos posibles caminos, como se ilustra en la Figura 63:

Parte común



A partir de la bifurcación



**Figura 63. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMG-07**

Un primer análisis consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apunta a que el ítem no presenta un problema de definición imprecisa del constructo pues las operaciones cognitivas anticipadas por las especialistas (como calcular el volumen de cada veladora a partir del conocimiento de sus aristas y posteriormente calcular mediante la división cuántas veces es mayor una de ellas con respecto a la otra) son coincidentes con lo que la especificación del ítem estipula que pretende evaluar: si el estudiante es capaz de calcular el volumen de un cubo.

Un análisis posterior, con información empírica obtenida a partir de los reportes verbales, corroboró que los estudiantes y docentes, al enfrentarse a esta tarea evaluativa evocan los conocimientos señalados en la especificación del reactivo y emplean, con ciertas variantes, las operaciones cognitivas anticipadas por las especialistas, donde las operaciones cognitivas medulares consisten en identificar que el problema implica una comparación de volúmenes y saber que la arista de un cubo toma la misma longitud para cada una de las tres dimensiones que lo componen.

La siguiente cita de uno de los alumnos entrevistados de NSE alto muestra cómo se evocan los conocimientos señalados en la especificación del reactivo, cuando después de que seleccionó la respuesta correcta se le pide que explique cómo llegó a ese resultado y verbaliza:

“tienes que poner atención en la arista y el volumen, porque puedes pensar cuántas veces es mayor el arista de la vela rosa comparada con la de la vela verde, pero no, te pide el volumen” (E07).

De la misma manera, uno de los alumnos de NSE medio verbaliza su comprensión de la arista cuando dice: “tiene forma de cubo entonces la arista sería como un lado de una cara...” (E05); posteriormente el alumno calcula los volúmenes de cada veladora y realiza dos procedimientos de solución: a) divide el volumen de la vela menor entre el volumen de la vela mayor; y b) multiplica el volumen de la vela menor por los valores de cada una de las opciones de respuesta hasta que obtiene como resultado el volumen de la vela mayor. Cuando la entrevista le pide que explique cómo llegó a ese resultado, responde:

“se supone que el cubo mide 8 centímetros cúbicos, el verde, entonces y el rosa serían  $6 \times 6 \times 6$ , serían 216... bueno, yo lo que traté era relacionar las respuestas pero puede haber otra forma, creo que sería dividiendo 216 entre 8...” (E05).



En conclusión, al comparar los procesos cognitivos que se manifestaron durante las entrevistas cognitivas con las habilidades y conocimientos que la especificación estipula evaluar así como las operaciones cognitivas anticipadas por las especialistas, no se encontró evidencia de que este ítem presente una definición imprecisa del constructo.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMG-07 así como los sondeos que se hicieron durante las entrevistas cognitivas, permitieron identificar una posible fuente de sesgo para ciertas subpoblaciones.

El docente que trabaja con alumnos de NSE bajo advirtió que podría ser una dificultad en algunos contextos que los alumnos se imaginen velas cuadradas, porque no son comunes, sino que ellos están acostumbrados a ver velas en forma de cilindro. Este mismo problema de sesgo potencial fue advertido por una de las especialistas en enseñanza de las matemáticas mediante la siguiente nota “El planteamiento del reactivo hace referencia a velas en forma de cubo. En muchas regiones los niños están familiarizados con velas o veladoras de formas cilíndricas y no son precisamente regulares. En otras regiones a las velas las llaman ceras o les dan otros nombres; por lo que este sería un elemento lingüístico que puede distraer a los alumnos para resolver el problema matemático; además, una vela de cúbica de 2 cm de arista no es plausible”

Vale decir que si bien estas advertencia sobre la forma cúbica de las velas convendría fuera considerada para, en su lugar, emplear otro tipo de objetos con forma de cubo, ninguno de los alumnos entrevistados ofreció evidencias de que incluir en el problema velas con forma de cubo fuese algo que los distrajera o confundiera; más bien, en aquellos casos en que no resolvieron de manera correcta el ítem, la información disponible indica que eso ocurrió porque no fueron capaces de comprender que el problema implicaba una comparación de volúmenes.

### ***Incomprensibilidad o problemas de legibilidad del ítem.***

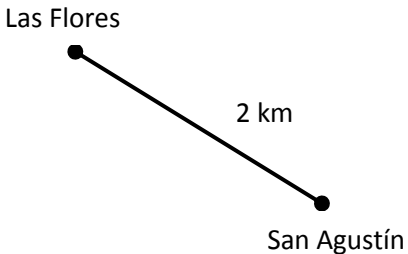
El análisis de las videgrabaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando ocurrió el caso de los alumnos seleccionaron una opción incorrecta, se debió principalmente a que no fueron capaces comprender que el problema planteaba una comparación de volúmenes o no fueron capaces de calcularlos.

Cuando a los entrevistados (alumnos y docentes) se les preguntó sobre la claridad en la redacción del ítem, o bien, sobre la inclusión de términos que les causaran confusión, todos respondieron que el reactivo les parecía claro en su formulación y había palabras que los confundieran.

### Anexo 36. Análisis de sesgo del reactivo SMH-11

El ítem SMH-11 (ver Figura 64) pertenece al eje de la asignatura de Matemáticas llamado Forma, espacio y medida; en el programa de Matemáticas corresponde al tema Formas geométricas y al subtema Figuras planas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de identificar las posibilidades de construcción (existencia y unicidad) de círculos.

Los pueblos de San Agustín y Las Flores se encuentran ubicados a 2 km de distancia entre sí.



¿Cuántas circunferencias de 4 km de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?

- Una
- Dos\*
- Más de dos
- Ninguna

\* Opción correcta

Figura 64. Reactivo SMH-11

En la primera etapa de este trabajo de investigación, el SMH-11 presentó un DIF uniforme que favorece a los alumnos con un nivel socioeconómico bajo con 0.78 lógitos con respecto a los alumnos de NSE alto; es decir, comparando alumnos con el mismo nivel de habilidad, el reactivo resultó 0.78 lógitos más fácil para aquéllos cuyo origen socioeconómico es Bajo en comparación con los estudiantes provenientes de un contexto socioeconómico favorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 19.

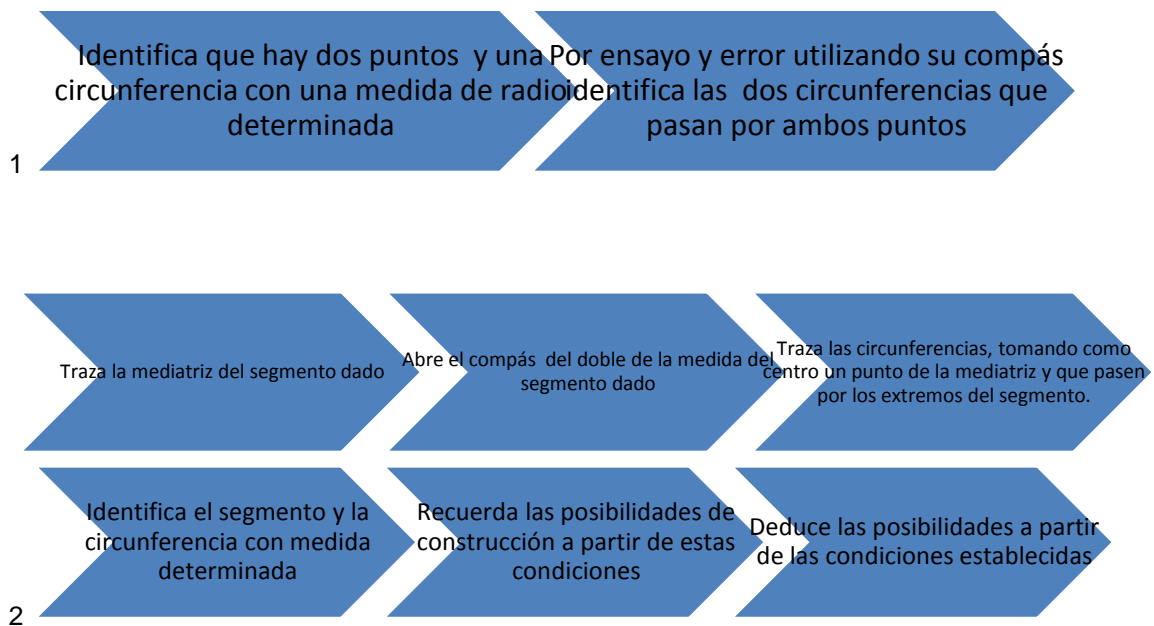
De los nueve entrevistados, ninguno de ellos respondió el reactivo de forma correcta; todos emplearon un procedimiento inadecuado cuando se enfrentaron a la tarea evaluativa que consistió en intentar identificar cuántas circunferencias podían pasar por los pueblos de San Agustín y Las Flores al mismo tiempo, y que dichas localidades fuesen los extremos de uno de sus diámetros; y justamente la comprensión errónea de lo que solicitaba el ítem, que probablemente se deba a un problema de legibilidad que se tratará más adelante, fue lo que desencadenó que los entrevistados utilizaran un procedimiento no apropiado a lo que se requería para solucionar la tarea evaluativa.

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis utilizadas en esta investigación y que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

#### ***Definición imprecisa del constructo***

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante disponga de los conocimientos sobre las condiciones de posibilidad y unicidad en la construcción de círculos y circunferencias dados algunos datos. Además se debe contar con los conocimientos y habilidades sobre construcción y trazo de figuras geométricas como mediatrices, rectas paralelas y rectas perpendiculares.

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían alguno de los dos senderos cognitivos que se ilustran en la Figura 65. Vale decir que tanto en el Excale como en la aplicación de las entrevistas los alumnos no utilizan su juego de geometría, por tanto los entrevistados dibujaron las circunferencias con medidas aproximadas:



**Figura 65. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMH-11**

Un primer análisis para identificar el potencial problema de definición imprecisa del constructo consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apunta a una coherencia entre ambas fuentes, en la que en ambas los elementos clave consiste en conocer las condiciones de posibilidad y unicidad en la construcción de círculos y circunferencias dados algunos datos y reconocer como mediatriz el segmento que separa las dos poblaciones que se incluyen en el problema.

Se llevó a cabo un segundo análisis a partir de los reportes verbales; sin embargo, debido a que ninguno de los entrevistados respondió de manera correcta este reactivo, no fue posible determinar la presencia o ausencia de una definición imprecisa del constructo como un posible problema de varianza irrelevante. Ciertamente los alumnos y docentes evocaron conocimientos distintos a los señalados por la especificación, sin embargo, al no haber encontrado la respuesta correcta, no se tienen la variación necesaria para distinguir si el hecho de que hayan evocado conocimientos diferentes a los que se pretendían medir fue lo que ocasionó que seleccionaran una opción incorrecta, o bien, si el reactivo está evaluando conocimientos distintos a los que busca medir.

***Ítem no accesible o sesgado***

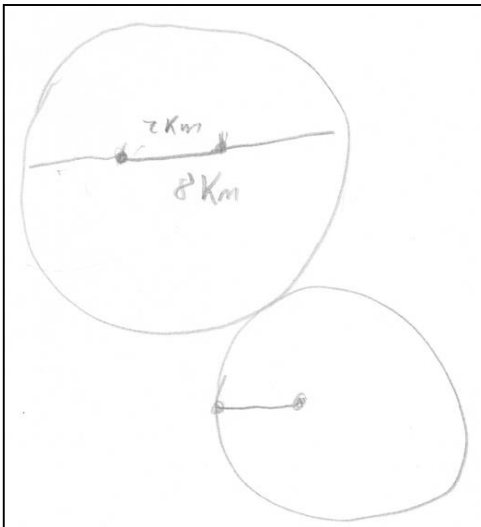
La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMH-11 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar alguna fuente de sesgo para los alumnos de NSE; de hecho, todos los entrevistados respondieron el ítem de forma errónea, pero que se sospecha ocurrió por un problema de legibilidad, el cual se tratará en el siguiente apartado (ver Tabla 49).

***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Lo que se halló es que cuatro de los alumnos (de los tres NSE) así como los tres docentes comprendieron erróneamente que el problema les planteaba encontrar cuántas circunferencias se pueden formar y que cumplieran con las siguientes condiciones: 1) que pasen por San Agustín y Las Flores al mismo tiempo; 2) que tengan 4 km de radio; y 3) que los dos pueblos sean los extremos de sus diámetros (ver Tabla 49). En realidad, el reactivo no pedía que se cumpliera la tercera condición; sin embargo, el hecho de que en la figura del pie del ítem se haya incluido una línea para ilustrar la distancia entre ambas localidades, al parecer fue lo que motivó esa confusión generalizada, de la cual hubo muchas evidencias.

La siguiente verbalización de uno de los alumnos de NSE alto ilustra el problema cuando explica porqué seleccionó la opción de ninguna circunferencia y dice:

“...porque si tengo 4 kilómetros de radio el diámetro sería de 8 entonces [señalando la línea de la figura que representa la distancia entre los dos pueblos, como se puede ver en la circunferencia superior de la siguiente figura que trazó] queda entre el círculo y queda pequeño entonces no alcanza a tomar [mientras traza en el aire una circunferencia]” (E08).



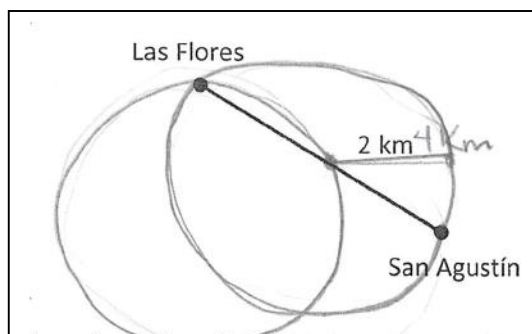
Luego cuando se le sondeó qué parte del reactivo fue lo que le hizo que pensara que la circunferencia debería tener el acomodo tal y como la dibujó en la circunferencia superior de la figura, el alumno responde:

“que dice ‘al mismo tiempo’... puede que el círculo esté cargado así [mientras dibuja la circunferencia inferior de la figura], pero dice ‘al mismo tiempo’ entonces aquí no... por lo tanto es ninguna” (E08).

Por su parte, la alumna de NSE bajo también parece comprender que los puntos deben ser los extremos del mismo diámetro, pues cuando selecciona la opción de respuesta d) que es “ninguna” y se le pregunta por qué eligió esa opción, responde:

“dice ‘un radio de 4’... entonces si hago un círculo y el radio es del centro hacia cualquiera de los lados de la superficie, y dice que de cuatro... pero la distancia es de 2, entonces sería más cortito...” (E02).

Parte del proceso que siguió la alumna de NSE medio para resolver el reactivo consistió en traza un círculo cuyos extremos del diámetro son la ubicación de los dos pueblos, como se muestra en la siguiente figura. La alumna señala que el círculo (derecho en la figura) mide 4 km de radio; vale decir que esta circunferencia no es plausible porque si tuviera un radio de 4 km, las dos localidades no podrían ser los extremos de su diámetro, tal y como ella la dibujó. Luego la entrevistada traza otro círculo (el izquierdo de la figura) en el que la circunferencia pasa por Las Flores y el punto medio entre los dos pueblos, y la alumna verbaliza: “pero éste ya no pasaría por San Agustín, entonces nada más sería una...[circunferencia]” (E04).



Nótese cómo en la figura anterior, las dos localidades se consideraron como extremos del diámetro de una de las circunferencias, y esa fue la confusión principal del reactivo, que se sospecha fue ocasionado por la inclusión de la línea que intenta ilustrar una distancia de 2 km. Así, cuando la entrevistadora le pregunta al docente que atiende alumnos de NSE alto “¿no puede haber una circunferencia de cuatro kilómetros que pase por esos dos puntos? [señalando los puntos que ubican a las dos localidades]”, el profesor responde:

“no, porque la distancia son dos... una circunferencia de cuatro viene siendo hasta acá [al tiempo que proyecta un punto siguiendo la misma línea que ilustra la distancia de los dos 2 km.]” (E09).

Además, otra muestra del problema de ilegibilidad del ítem es que al menos cinco entrevistados tuvieron dificultades para comprender la pregunta del pie del ítem. Por ejemplo, uno de los alumnos de NSE bajo señala que el pie del reactivo le parece confuso, y cuando se le pide que explique qué es lo que le causa confusión, verbaliza la pregunta del pie del ítem: “¿Cuántas circunferencias de 4 kilómetros de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?” (E01).

Otro de los alumnos entrevistados (NSE alto) después de leer en una primera ocasión el pie del reactivo, dice:

“ok, aquí lo voy a leer otra vez porque no entendí muy bien a la pregunta, esto sí está muy claro [refiriéndose a la figura]... no sé muy bien si al decir circunferencia se refiere a que sean círculos, cuántos círculos pasan así [mientras con su lápiz forma círculos imaginarios sobre la figura del pie del ítem, de tal forma que San Agustín y Las Flores sean extremos de su diámetro], o si del círculo, que sea como un hilo que se estire, cuántas veces son...” (E07).

Este mismo alumno comprende que la línea del pie del ítem que ilustra la distancia de 2 km entre los dos pueblos debe ser parte del diámetro de las posibles circunferencias, y que por tanto ninguna circunferencia pasa por ambas localidades, cuando verbaliza:



“ok, entonces tengo que pensarlo de esta manera... de cualquier manera no cabría, porque ya cuatro kilómetros de radio ya sería el doble de esto [mientras simula una línea de 4 km sobre la línea de la figura]... no cabe ningún círculo ahí” (E07).

Uno de los docentes (de NSE alto) lee en tres ocasiones el pie del reactivo, en especial la pregunta “¿Cuántas circunferencias de 4 km de radio pueden pasar por San Agustín y por Las Flores al mismo tiempo?” y verbaliza:

“eso sí no entiendo... no le entiendo... aquí son dos kilómetros [señalando la línea de la figura del pie del ítem que ilustra la distancia entre los dos pueblos]... no, no es comprensible la pregunta” (E09).

Cuando la entrevistadora le sondea porqué no se comprende, el docente responde:

“porque este dato [señalando la línea de la figura del pie del ítem que ilustra la distancia entre los dos pueblos] no sé qué tiene que ver con los 4 kilómetros de radio de la circunferencia” (E09).

**Tabla 49. Cantidad de entrevistados según el tipo de problema de varianza irrelevante al constructo que presentó el ítem SMH-11, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Tipo de problema que presenta el reactivo			
				Definición imprecisa del constructo	Ítem no accesible o sesgado	Procedimientos e instrucciones complejas, poco claras e intuitivas	Incomprensibilidad o problemas de legibilidad del ítem
¿El entrevistado respondió correctamente el ítem?	No (N = 9)	NSE	Bajo	0	0	0	2
			Medio	0	0	0	2
			Alto	0	0	0	3
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>7</b>
	Sí (N = 0)	NSE	Bajo	0	0	0	0
			Medio	0	0	0	0
			Alto	0	0	0	0
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Total</b>				<b>0</b>	<b>0</b>	<b>0</b>	<b>7</b>

**Notas:** NSE=Nivel Socioeconómico  
**Fuente:** elaboración propia

### Anexo 37. Análisis de sesgo del reactivo SMA-05

El ítem SMA-05 (ver Figura 66) pertenece al eje de la asignatura de Matemáticas llamado Sentido numérico y pensamiento algebraico; en el programa de Matemáticas corresponde al tema y subtema del mismo nombre: Problemas aditivos. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de resolver problemas aditivos con números fraccionarios con distinto denominador.

Valentina estuvo leyendo un libro. El viernes leyó  $\frac{3}{8}$  partes, el sábado  $\frac{1}{4}$  parte y el domingo  $\frac{5}{16}$  partes. ¿Qué parte del libro le faltó leer?

$\frac{15}{512}$

$\frac{15}{16}$

$\frac{9}{28}$

$\frac{1}{16}$  \*

\* Opción correcta

Figura 66. Reactivo SMA-05

En la primera etapa de este trabajo de investigación, el SMA-05 presentó un DIF no uniforme; entre individuos del primer tercil de habilidad matemática, este ítem es 0.73 lógitos más fácil para los alumnos de NSE bajo, en cambio, entre individuos del tercil más alto de habilidad matemática, este reactivo es 0.81 lógitos más fácil para los alumnos de NSE alto.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 20.

De los seis estudiantes que respondieron el reactivo, la mitad lo contestó de manera correcta (los dos alumnos de NSE alto y uno de NSE medio), y los tres restantes, de NSE bajo y uno de NSE medio, seleccionaron alguna de las opciones incorrectas. Los tres docentes respondieron el reactivo de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, lo que se encontró es que de los tres alumnos que respondieron el ítem de forma correcta, dos de ellos utilizaron algún procedimiento de solución adecuado, tal como realizar la suma de las fracciones  $\frac{3}{8} + \frac{1}{4} + \frac{5}{16}$  convirtiendo las fracciones  $\frac{3}{8}$  y  $\frac{1}{4}$  en sus equivalentes con denominador 16, de tal manera que  $\frac{6}{16} + \frac{4}{16} + \frac{5}{16} = \frac{15}{16}$  y el resultado se lo restaron a la unidad. El otro alumno que seleccionó la opción correcta, aparentemente adivinó, como se explicará párrafos más adelante.

Por su parte, los alumnos que respondieron de manera incorrecta, utilizaron algún procedimiento inapropiado o incompleto para su solución. Inapropiado, porque no supieron sumar las fracciones con distinto denominador y realizaron una suma de numeradores (3+1+5) y posteriormente una suma de denominadores (8+4+16), obteniendo el siguiente resultado 9/28; hubo otros alumnos que en lugar de sumar, realizaron una multiplicación o una división de las tres fracciones, y uno de ellos incluso convirtió las fracciones en decimales. Incompleto, porque si bien realizaron correctamente la suma de fracciones con distinto denominador, no realizaron la resta a la unidad, que representa a todo el libro que leyó Valentina (ver Tabla 50).

Vale señalar que los tres docentes utilizaron un procedimiento de solución adecuado. Para realizar la suma de fracciones, dos de ellos siguieron la ruta en la que identificaron al 16 como múltiplo de los denominadores 4 y 8; en cambio, uno de los profesores, calculó el mínimo común múltiplo de los denominadores 4, 8 y 16 y a partir de ahí generó las fracciones equivalentes con mismo denominador.

**Tabla 50. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMA-05, desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 3)	Nivel Socioeconómico	Bajo	2	0	0	--
			Medio	1	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>
	Sí (N = 3)	Nivel Socioeconómico	Bajo	--	0	0	0
			Medio	--	1	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>2</b>
	<b>Total</b>			<b>3</b>	<b>1</b>	<b>0</b>	<b>2</b>
	<b>Fuente:</b> elaboración propia						

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente, los cuales se desglosan para cada uno de los potenciales problemas de sesgo que se indagaron: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem.

***Definición imprecisa del constructo***

De acuerdo con la especificación del ítem, para resolverlo correctamente se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Otorgar sentido a cantidades expresadas como fracción.
- Interpretar cantidades fraccionarias en contextos de medición.
- Expresar cantidades utilizando fracciones.
- Emplear los algoritmos de la adición, sustracción de números naturales mediante equivalencias de fracciones.
- Resolver, con base en la equivalencia de fracciones, problemas aditivos simples y combinados, directos e inversos que involucren números fraccionarios.

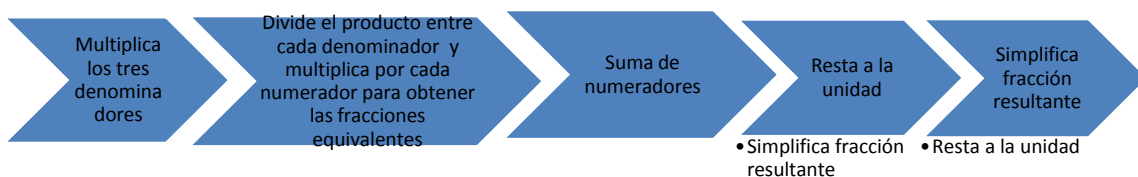
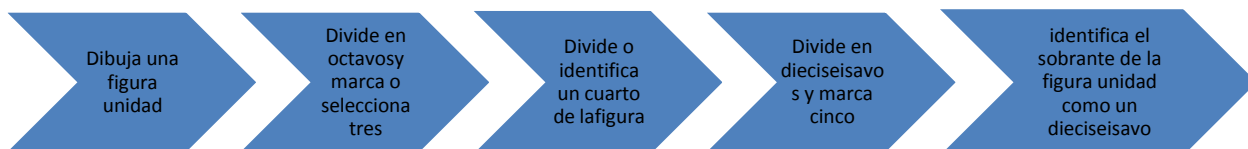
- Identificar y producir familias de fracciones equivalentes mediante la amplificación y simplificación, reconociendo que todas ellas expresan una misma cantidad

Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes emplearían alguno de los cuatro siguientes procesos cognitivos, que en un inicio comparten subprocesos en común (ver Figura 67):

- Parte común



- A partir de la bifurcación



**Figura 67. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMA-05**

Al contrastar las habilidades y los conocimientos estipulados en la especificación del ítem con los procesos cognitivos que anticiparon las especialistas, se encontró que existe coherencia entre ambos. Por ejemplo, para la habilidad “Emplear los algoritmos de la adición, sustracción de números naturales mediante equivalencias de fracciones”, las especialistas anticiparon en cada uno de los procesos, que los sustentantes deberían hacer una suma con fracciones de distinto denominador para conocer la proporción del libro que se ha leído; el proceso a emplear para realizar dicha suma podría variar, sin embargo, todos requeriría poner en juego su habilidad para trabajar con equivalencias de fracciones, desde la utilización del mínimo común múltiplo como denominador común hasta la identificación de que 16 es múltiplo del 4 y 8, entre otros.

Ahora bien, los análisis de las entrevistas cognitivas sugieren que los entrevistados (estudiantes y docentes), al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con algunas variantes, las operaciones cognitivas anticipadas por las especialistas. En otras palabras, la evidencia disponible muestra que cuando los estudiantes responden al reactivo tienen que: identificar que el problema involucra suma y resta de fracciones con distinto denominador; convertir las fracciones en equivalentes con un mismo denominador; realizar la suma de fracciones; y el resultado de la suma, restárselo a la unidad.

La siguiente cita de uno de los alumnos que respondió correctamente al reactivo ilustra que el proceso que empleó evoca los conocimientos y habilidades que están contempladas en la especificación del reactivo:

“Tenía nada más que encontrar lo que le faltó leer, pero para eso tengo que saber qué estuvo leyendo... pero como está en octavos, cuartos y dieciseisavos tenía que ponerlos todos parejos para poder sumarlos...”  
(E08).

En resumen, la contrastación entre los reportes verbales con los procesos cognitivos anticipados por las especialistas y a su vez la comparación con la especificación del reactivo apunta que este ítem no presenta el problema de sesgo sobre una definición imprecisa del constructo.

### ***Ítem no accesible o sesgado***

Una de las especialistas en enseñanza de las matemáticas anticipó una posible fuente de sesgo en el pie del reactivo, pues “Los alumnos identifican la unidad como el todo y un libro tiene algunas partes que por costumbre los niños no ‘leen’ como las hojas legales, los índices, etc. Por lo que al preguntar qué parte leen pueden realizar diferentes interpretaciones de cuál es el todo”; sin embargo, los

análisis de las respuestas de los entrevistados, no permiten sostener que exista esa confusión, más bien, lo que fue patente a partir de las entrevistas es lo indispensable que resulta para resolver correctamente el ítem el hecho de considerar en el proceso cognitivo al libro como la unidad a partir de la cual se le restaría la parte ya leída.

Dos de los docentes (que trabajan con alumnos de NSE bajo y medio) señalaron que si bien el reactivo en su formulación le parece claro, no obstante, anticiparon que posiblemente a algunos de sus alumnos se les dificultaría identificar cuál es el entero, que en el ítem corresponde al libro que leyó Valentina. De hecho, los dos alumnos que respondieron correctamente el reactivo sin haber adivinado, fueron los únicos que de forma explícita señalaron que para la resolución del problema se requeriría que el total de páginas del libro se tomara como el entero y que al restarle la suma del total de páginas, se obtendría la cantidad de páginas que le faltaría por leer.

En síntesis, no se encontró evidencia de que el ítem presente algún sesgo o sea inaccesible para alguno de los tres grupos de alumnos que se exploraron en este trabajo; más bien, la identificación del libro como un entero es parte del proceso cognitivo que es fundamental para resolver el reactivo, así como la capacidad del sustentante para sumar fracciones con distinto denominador. Aquellos alumnos que tuvieron dificultades para realizar alguno de estos dos subprocesos, son los que no lograron resolver correctamente el ítem; y eso lo que parece explicar que unos alumnos hayan elegido una opción incorrecta, o bien, declarado su desconocimiento para llevar a cabo el procedimiento que requería la tarea evaluativa.

Una de las alumnas que seleccionó la opción correcta, al parecer adivinó. Mientras pensaba en voz alta para seleccionar la respuesta correcta, dice:

“creo yo que... puede ser la opción número uno y la cuatro, una dieciseisava parte del libro, o la opción “a” de quince... quince quinientosdoceava parte del libro... entonces me voy por la d...” (E04).

Además, esa alumna tiene comprensiones erróneas o desconocimiento de elementos clave para responder correctamente el reactivo, tales como: 1) no considera que el libro es la unidad; señala que el ítem no le resulta del todo claro porque le hace falta que le digan la cantidad total de páginas que tenía el libro, como se ilustra en su respuesta cuando se le preguntó porqué no le resultó claro el reactivo, responde: “porque no me dio el número exacto... o un acercamiento a las páginas que tiene el libro, entonces yo no podría saber cuánto es lo que le falta

por leer porque no me dio un número aproximado a las páginas del libro” (E04); y 2) considera que las fracciones son valores aproximados.

***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando seleccionaban una opción incorrecta se debía principalmente a su incapacidad para realizar al menos uno de los dos siguientes procesos que resultan fundamentales para resolver el reactivo: la suma de fracciones con distinto denominador y la identificación del libro como la unidad.

Evidencias de que hubo una comprensión adecuada del pie del ítem se mostró en todos los entrevistados.

Por ejemplo, cuando a la alumna que aparentemente adivinó se le preguntó si entendía lo que tenía que resolver, dice:

...”mh..mh.. [al tiempo que asiente con la cabeza]... encontrar lo que había leído... cuánto es lo que le faltaba leer...” (E04).

Otro alumno después de leer el pie del reactivo y verbalizar lo que comprendió, afirma:

“al ver el problema ya sé que... el libro, el total de páginas se tomará como un entero... de todo el libro... entonces aquí lo que haría sería sumarlas todas y luego al entero le resto las [páginas] que leyó... y lo que salga será lo que faltó por leer” (E07).



### Anexo 38. Análisis de sesgo del reactivo SMG-05

El ítem SMG-05 (ver Figura 68) pertenece al eje de la asignatura de Matemáticas llamado Sentido numérico y pensamiento algebraico; en el programa de Matemáticas corresponde al tema Significado y uso de las literales y al subtema Patrones y fórmulas. El reactivo pertenece a la especificación del Excale que busca evaluar si los sustentantes son capaces de calcular el valor de un término dado de una sucesión de números con signo.

<p>Considera que la siguiente sucesión numérica continúa.</p> <p style="text-align: center;"><math>-3, 2, 7, 12, 17, \dots</math></p> <p>¿Cuál será el valor del término que ocupe el lugar 35 en la sucesión?</p> <p><input type="radio"/> 27</p> <p><input type="radio"/> 119</p> <p><input type="radio"/> 167*</p> <p><input type="radio"/> 175</p> <p>* Opción correcta</p>
---

Figura 68. Reactivo SMG-05

En la primera etapa de este trabajo de investigación, el SMG-05 presentó un DIF no uniforme que en el primer tercil de habilidad matemática favorece a los alumnos con un nivel socioeconómico alto con 0.69 lógitos y en el tercil más alto de habilidad favorece con 0.72 lógitos a los alumnos de NSE bajo; es decir, comparando alumnos con bajo nivel de habilidad matemática, el reactivo resultó 0.69 lógitos más fácil para aquéllos cuyo origen socioeconómico es Alto; en cambio, haciendo la misma comparación pero entre los alumnos que demostraron un nivel más alto de habilidad matemática, el ítem fue 0.72 lógitos más fácil para los estudiantes provenientes de un contexto socioeconómico desfavorecido.

El análisis de la información recolectada de cada uno de los nueve entrevistados, en lo que concierne a este reactivo se incluye en el E-Anexo 21.

De los seis estudiantes que respondieron el reactivo, la mitad lo respondió de forma correcta (dos alumnos de NSE alto y el alumno de NSE bajo considerado por su profesor como de bajo aprovechamiento), mientras que los dos estudiantes de NSE medio y la alumna de NSE bajo considerada por su profesor de buen rendimiento, lo respondieron de forma errónea. Por su parte, los tres docentes lo respondieron de forma adecuada.

Respecto a la adecuación del procedimiento empleado para enfrentarse a la tarea evaluativa, los tres alumnos que lo respondieron de forma correcta utilizaron un procedimiento adecuado (ver Tabla 51), en el que: 1) identificaron que el problema implicaba una sucesión numérica creciente; 2) descubrieron que entre el valor de cada posición había un incremento de 5 unidades; 3) dedujeron la fórmula  $5n-8$  que sirve para encontrar el valor de una posición dada; 4) comprobaron la fórmula con alguna de las primeras cinco posiciones de la secuencia, las cuales están dadas en el pie del reactivo; 6) operaron la fórmula  $5n-8$  para obtener el valor de la posición 35 de la siguiente manera  $5(35)-8=175-8=167$ ; 7) seleccionaron la opción c)

Por su parte, los dos alumnos de NSE medio y que seleccionaron una opción incorrecta utilizó un procedimiento de solución incorrecto, en el que intentaron a partir de aproximaciones o estimaciones obtener el resultado correcto. Por otro lado, la alumna de NSE bajo que seleccionó una opción incorrecta, si bien encontró en la secuencia que el incremento constante era de 5 unidades, hubo algunos elementos del proceso de solución que desconocía, por lo que no consideró que el valor de la primera posición no iniciaba en 5 y así multiplicó el incremento entre valores (5) de la secuencia por la posición que se buscaba conocer (35),  $35 \times 5 = 175$  obteniendo una respuesta errónea (ver Tabla 51).

**Tabla 51. Cantidad de alumnos según la adecuación del procedimiento empleado para responder el reactivo SMG-05 desglosado por nivel socioeconómico y si respondió correctamente al ítem**

				Proceso de solución del problema			
				Utilizó un procedimiento incorrecto en la resolución del problema	Aparentemente adivinó	Hubo algunos elementos del proceso de solución que desconocía	Utilizó un procedimiento de solución de manera correcta
¿El alumno respondió correctamente el ítem?	No (N = 3)	Nivel Socioeconómico	Bajo	0	0	1	--
			Medio	2	0	0	--
			Alto	0	0	0	--
			<b>Subtotal</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>0</b>
	Sí (N = 3)	Nivel Socioeconómico	Bajo	--	0	1	0
			Medio	--	0	0	0
			Alto	--	0	0	2
			<b>Subtotal</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>2</b>
<b>Total</b>			<b>2</b>	<b>0</b>	<b>2</b>	<b>2</b>	

**Fuente:** elaboración propia

A continuación se presentan los resultados de los análisis de sesgo realizados a partir de las técnicas de pensamiento en voz alta y sondeo concurrente. Se desglosan según las categorías de análisis que corresponden con tres de los Elementos del Diseño Universal de las Evaluaciones: 1) Definición imprecisa del constructo; 2) Ítem no accesible o sesgado; y 3) Incomprensibilidad o problemas de legibilidad del ítem. Vale señalar que también se analizó el potencial problema denominado *Procedimientos e instrucciones complejas, poco claras e intuitivas*; sin embargo, se omite porque no se encontraron evidencias de su presencia.

### **Definición imprecisa del constructo**

Para resolver el reactivo, de acuerdo con su especificación, se requiere que el sustentante domine las siguientes habilidades y tenga los siguientes conocimientos previos:

- Suma, resta y multiplicación de números con signo.
- Calcular el valor numérico de expresiones lineales de la forma  $ax + b$  o  $ax - b$ .
- Habilidad para encontrar semejanzas y diferencias en conjuntos de datos.
- Generalizar procedimientos a partir del análisis de casos particulares.

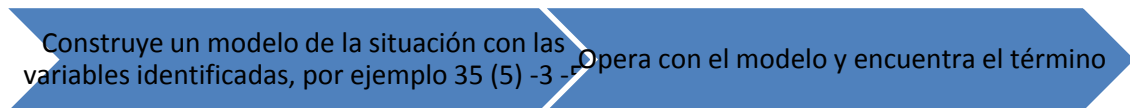
Por su parte, las especialistas en enseñanza de las matemáticas anticiparon que los sustentantes utilizarían un sendero cognitivo que tiene una parte común y

luego podría tomar alguno de los tres posibles caminos, como se ilustra en la Figura 69:

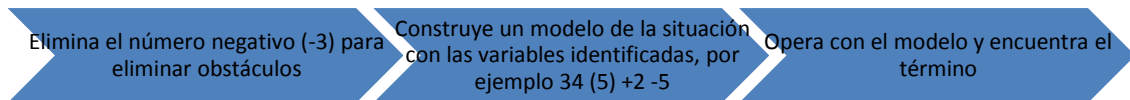
Parte común



A partir de la trifurcación



Antes de encontrar el término, pone a prueba su modelo con un número menor para verificar.



Antes de encontrar el término, pone a prueba su modelo con un número menor para verificar.

**Figura 69. Proceso(s) cognitivo(s) anticipados por las especialistas en enseñanza de las matemáticas para el reactivo SMG-05**

Un primer análisis para identificar el potencial problema de definición imprecisa del constructo consistió en contrastar entre los conocimientos y habilidades estipulados en la especificación con los procesos cognitivos anticipados por las especialistas. Este primer análisis apuntó existe coherencia entre la especificación y los procesos anticipados por las especialistas, en el que aspecto central consiste en encontrar un patrón en la serie numérica o elaborar una fórmula que describa dicho patrón.

Un análisis posterior y realizado a partir de los reportes verbales, sugiere que los entrevistados (estudiantes y docentes) al enfrentarse a esta tarea evaluativa, evocan los conocimientos señalados en la especificación del reactivo y emplean, con ciertas variantes, las operaciones cognitivas anticipadas por las especialistas, donde la operación cognitiva medular consiste en deducir la fórmula para encontrar cualquier valor en la sucesión, o bien, encontrar el patrón en el los valores a partir de la segunda posición terminan en 2 o en 7.

La siguiente cita de uno de los alumnos entrevistados muestra cómo se evocan los conocimientos señalados en la especificación del reactivo, cuando después de que seleccionó la respuesta correcta se le pide que explique cómo llegó a ese resultado y describe:

“primero de cuánto en cuánto iba la sucesión, entonces vi que iba de cinco, pero la sucesión no empieza en el cinco, empieza en -3; [la diferencia] entre 5 y -3 es 8... y para nada más comprobar si estaba bien lo multipliqué primero por el término 1 que es éste y ya me dio -3, entonces sabía que estaba bien esta fórmula y ya me fui a la [posición] 35...” (E08).

En el mismo sentido, un alumno de NSE bajo resuelve el problema utilizando una estrategia distinta, en la que no emplea una fórmula, sino que además de identificar que el patrón de la sucesión se conforma a partir de incrementar 5 unidades, luego continúa sumando 5 unidades al último término y cuando llega a la 7ª posición de la sucesión, la cual tiene un valor de 27, descarta la primera opción de respuesta; posteriormente continúa calculando los valores de la sucesión hasta la 15ª posición e identifica un patrón: que el último dígito de los valores de la sucesión es 2 o 7, a partir de ello infiere que el resultado es 167, el que está ubicado en la opción c). Descarta la opción b) y d) porque sus valores no terminan en 2 o 7.

A partir de la evidencia recolectada no hay elementos que indiquen que este reactivo tienen una definición imprecisa del constructo, sino que, como ya se señaló, para responderlo de forma correcta los entrevistados pusieron en juego los conocimiento y habilidades estipulados en la especificación del SMG-05.

### ***Ítem no accesible o sesgado***

La evidencia recolectada mediante el pensamiento en voz alta mientras los entrevistados se enfrentaban al reactivo SMG-05 así como los sondeos que se hicieron durante las entrevistas cognitivas, no permitieron identificar elementos que pudieran favorecer u obstaculizar a algunas de las subpoblaciones según su NSE. Si se considera que durante la primera etapa de esta investigación este

reactivo tuvo un DIF no uniforme que en el primer tercil de habilidad matemática es más difícil para los alumnos de NSE bajo y en el tercil más alto es más difícil para los alumnos de NSE alto, sin embargo, en la segunda etapa de este trabajo los alumnos tanto de NSE alto como NSE bajo y con bajo rendimiento según la visión de sus profesores, lo respondieron de forma correcta, además, no se encontraron elementos que hagan pensar que este reactivo pueda discriminar favorable o desfavorablemente por el NSE de los sustentantes, sino que para este reactivo lo que hizo la diferencia fue la comprensión a cabalidad de los conocimientos y habilidades establecidos en la especificación del reactivo.

***Incomprensibilidad o problemas de legibilidad del ítem.***

El análisis de las videograbaciones de las entrevistas cognitivas no aporta elementos para afirmar sobre un posible problema de incomprensibilidad o legibilidad del ítem. Más bien, lo que se halló es que tanto los alumnos como los docentes entrevistados comprendían el pie del reactivo y que cuando ocurrió el caso de los alumnos seleccionaron una opción incorrecta, se debió principalmente a que no fueron capaz de identificar alguno de los dos patrones en la secuencia: que los valores de cada posición estaban dados por la fórmula  $5n-8$ , o bien, que el último dígito de los valores positivos podría ser sólo 2 ó 7.

Cuando a los entrevistados (alumnos y docentes) se les preguntó sobre la claridad en la redacción del ítem, o bien, sobre la inclusión de términos que les causaran confusión, todos respondieron que el reactivo les parecía claro en su formulación y había palabras que los confundieran.