



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

FACULTAD DE EDUCACIÓN

ANÁLISIS DE FUNCIONAMIENTO DIFERENCIAL DE ÍTEMS EN
EL PORTAFOLIO DEL SISTEMA DE EVALUACIÓN DEL
DESEMPEÑO PROFESIONAL DOCENTE DEL AÑO 2013

Por

MIGUEL ÁNGEL PLAZA MUÑOZ

Tesis presentada a la Facultad de
Educación de la Pontificia
Universidad Católica de Chile,
para optar al grado académico de
Magister en Educación con mención en
Evaluación de Aprendizajes

Profesora Guía: Marianela Navarro Ciudad

Agosto 2017
Santiago, Chile

© 2017, Miguel Ángel Plaza Muñoz

© 2017, Miguel Ángel Plaza Muñoz

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica que acredita al trabajo y a su autor.

*A María Luisa, compañera de vida que compartió las
luces y sombras de este trabajo.*

*A mis hijas Camila y Constanza, y a mi hijo Salvador,
cuya sola existencia da sentido a la vida, y es razón
para luchar a diario por un mundo mejor.*

AGRADECIMIENTOS

Por hacer posible que haya logrado dar este importante paso, agradezco: en primer lugar, a mi familia, especialmente a mi gran compañera María Luisa, y mis hijas e hijo: Camila, Constanza y Salvador, por su infinito amor, apoyo y paciencia, durante todas las horas robadas a un tiempo que les pertenecía, para llevar a cabo este trabajo.

A mi madre Gabriela Muñoz Córdova, educadora que, sin saberlo, inculcó esta vocación por aprender y enseñar. A mi hermano Camilo y mi hermana Gabriela por el cariño y apoyo permanente.

A mi profesora guía, Sra. Marianela Navarro, por su comprensión, apoyo y paciencia.

Al Sr. José Giménez, y a la Sra. Marcela Cazaux, dos grandes maestros que en su posición de jefes, siempre tuvieron la mejor disposición para darme las condiciones laborales que permitieron comenzar el desafío de este Magister.

A mis amigos Francisco Martínez-Conde y Benjamín Espinace, por su apoyo en la comprensión de los procedimientos estadísticos. Y a Osvaldo Carvajal por su ayuda en la edición de este trabajo.

A los académicos de la Pontificia Universidad Católica, especialmente a la profesora María Verónica Santelices, quien fue un apoyo y gran ayuda en el desarrollo de esta tesis, tanto en la elección del tema, como en la búsqueda de soluciones ante las dificultades metodológicas que se presentaron.

Al Sr. Ángel Fidalgo, por su actitud siempre dispuesta a colaborar con este trabajo, a través de orientaciones metodológicas, además de facilitar y autorizar el uso del software de su autoría GMHDIF.

Al Sr. Lorenzo Urbano, por facilitar y autorizar el uso del software *Factor*

Al Sr. Sergio Domínguez, por facilitar y autorizar el uso de una herramienta informática de su autoría, para el cálculo del estadístico Alpha Ordinal.

A la Comisión Nacional de Investigación Científica y Tecnológica, por el apoyo brindado a través de la Beca de Magister en Chile, para profesionales de la Educación.

ÍNDICE DE TABLAS

Tabla 1.....	59
Tabla 2.....	62
Tabla 3.....	67
Tabla 4.....	68
Tabla 5.....	70
Tabla 6.....	73
Tabla 7.....	74
Tabla 8.....	74
Tabla 9.....	77
Tabla 10.....	78
Tabla 11.....	79
Tabla 12.....	80
Tabla 13.....	84
Tabla 14.....	85
Tabla 15.....	86
Tabla 16.....	86
Tabla 17.....	87
Tabla 18.....	88
Tabla 19.....	88
Tabla 20.....	89
Tabla 21.....	90
Tabla 22.....	92
Tabla 23.....	94
Tabla 24.....	94
Tabla 25.....	95
Tabla 26.....	96
Tabla 27.....	97
Tabla 28.....	98
Tabla 29.....	99
Tabla 30.....	100
Tabla 31.....	101

ÍNDICE DE FIGURAS

Figura n°1.....	18
Figura n°2.....	22
Figura n°3.....	29
Figura n°4.....	48
Figura n°5.....	49
Figura n°6.....	51
Figura n°7.....	52
Figura n°8.....	58
Figura n°9.....	102

Tabla de contenido

AGRADECIMIENTOS.....	iii
ÍNDICE DE TABLAS.....	iv
ÍNDICE DE FIGURAS.....	v
RESUMEN.....	ix
I INTRODUCCIÓN.....	1
1.1 Antecedentes.....	2
1.2 Delimitación del problema de investigación.....	11
1.3. Objetivos.....	13
1.3.1 Objetivo General.....	13
1.3.2 Objetivos Específicos.....	13
II MARCO DE REFERENCIA.....	14
2.1 Medición de la Calidad Docente.....	14
2.2 Evaluación del Desempeño Profesional Docente en Chile.....	19
2.2.1 Descripción del portafolio.....	21
2.3 Funcionamiento Diferencial de Ítems.....	25
2.3.1 Tipos de DIF.....	27
2.3.2 Métodos de detección del DIF.....	27
2.4 Revisión de los resultados de la Evaluación Docente y definición de variables para el análisis.....	30
2.4.1 Sexo.....	31
2.4.2 Edad.....	32
2.4.3 Zona Geográfica (rural/urbana).....	33
2.4.4 Nivel de Enseñanza y Asignatura.....	36
III Marco Metodológico.....	40
3.1 Población y muestras.....	40
3.2.1 Base de datos y su preparación.....	41
3.2.2 Extracción de Muestras.....	42
3.3 Consideraciones metodológicas acerca del instrumento de medición.....	44
3.4 Operacionalización de las variables en estudio.....	47
3.4.1 Sexo.....	47
3.4.2 Zona geográfica (Rural/Urbano):.....	47

3.4.3 Edad (por rangos):	48
3.4.4 Nivel de Enseñanza:	49
3.4.5 Asignatura:	50
3.5 Procedimientos de análisis de datos	53
3.5.1 Análisis descriptivos e inferenciales	53
3.5.2 Análisis de Impacto	54
3.5.3 Análisis DIF	55
2.2.3 Estadísticos Generalizados de Mantel-Haenszel	57
2.2.4 Diferencia de Medias Estandarizadas	62
IV RESULTADOS.....	64
4.1 Análisis Inferenciales:	64
4.2 Resultados por variables	66
4.2.1 Sexo	66
4.2.2 Zona Geográfica	68
4.2.3 Edad	69
4.2.4 Nivel de Enseñanza	71
4.2.5 Asignatura.....	80
4.3 Resumen por indicador	102
V DISCUSIÓN Y CONCLUSIONES.....	103
5.1 Discusión	103
5.2 Conclusiones.....	108
5.2.1 Limitaciones y Proyecciones	111
REFERENCIAS.....	114
ANEXO 1: Análisis descriptivos e inferenciales	126
1 Análisis de valores perdidos	127
2 Frecuencias por variables	127
3 Extracción de Componentes Principales.....	129
4 Pruebas inferenciales por variables.....	130
4.1 Prueba T de Student variable Sexo.....	130
4.2 Prueba T de Student variable Zona Geográfica.....	130
4.3 ANOVA variable Edad	131

4.4 ANOVA variable Nivel de Enseñanza.....	132
4.5 ANOVA variable Asignatura (Segundo Ciclo Básico).....	134
4.6 ANOVA variable Asignatura (Educación Media).....	136
ANEXO 2: Análisis de Impacto y DIF.....	138
1. ANÁLISIS IMPACTO Y DIF POR VARIABLE.....	139
1.1 SEXO.....	139
1.2 ZONA GEOGRÁFICA.....	140
1.4 NIVEL DE ENSEÑANZA.....	145
1.5 ASIGNATURA.....	153

RESUMEN

El presente estudio analiza la presencia de Funcionamiento Diferencial de Ítems (DIF) en los indicadores que conforman el instrumento *portafolio* de la Evaluación Docente aplicada a los profesores del sistema escolar municipal chileno. Específicamente, se indaga en aquellas diferencias que, siendo significativas para diferentes grupos presentes en una población heterogénea de profesores, con el fin de determinar si estas pudieran ser atribuidas a las características psicométricas del instrumento *portafolio*. Los análisis abarcan 5 variables: Sexo, Zona Geográfica (rural o urbana), Edad (por rangos), Nivel de enseñanza en el cual se desempeña el docente evaluado y la Asignatura que imparte.

Se aplican dos procedimientos de análisis estadístico, adaptando las condiciones de implementación, de acuerdo con las características específicas de cada una de las variables estudiadas. Las conclusiones fueron obtenidas mediante la aplicación de un criterio de convergencia de resultados entre los resultados arrojados por el estadístico generalizado de Mantel-Haenszel (Ordinal) y un índice obtenido a partir del procedimiento de Diferencia de Medias Estandarizadas (SMD).

Los análisis se aplicaron a cada uno de los indicadores, sobre 33 diferentes comparaciones entre grupos focales y de referencia definidos por cada una de las variables ya mencionadas, pudiéndose observar que todos los indicadores presentan DIF severo, en al menos 3 comparaciones, siendo el indicador 6.1: *Calidad del inicio de la clase*, el que concentra una mayor cantidad de análisis con presencia de DIF severo (18), siempre contra los grupos focales respectivos. Del mismo modo, este indicador es el único que presenta DIF para las variables demográficas *Sexo* y *Edad*, en afectando a los grupos “Hombres” y “Mayores de 61 años”, respectivamente. Por otra parte, al observar el comportamiento del indicador 7.4: *Indicador por asignatura o área* se aprecia que presenta DIF severo en 12 comparaciones, 11 de las cuales corresponden a

funcionamiento diferencial en contra del grupo de referencia, lo que implica que, pese a que los docentes pertenecientes a estos grupos obtienen un desempeño global superior en el portafolio, se ven desfavorecidos frente a los grupos focales definidos para las variables *Nivel de Enseñanza y Asignatura*.

Observando los resultados arrojados por los múltiples análisis implementados, se concluye, en primer lugar, la necesidad de revisar la pauta que se aplica para la medición del indicador 6.1: *Calidad del inicio de la clase*. Del mismo modo, dado que los análisis aplicados sobre variables relacionadas con aspectos de la formación, desarrollo y/o ejercicio profesional (*Nivel de Enseñanza y Asignatura*), arrojaron la presencia de DIF, especialmente en los indicadores 1.2: *Relación entre actividades y objetivos* y 7.4: *Indicador por asignatura* y, dado que el análisis DIF no informa respecto de las causas del mismo, se hace necesario implementar nuevos estudios tendientes a encontrar las causas de este DIF, indagando en variables como la dificultad del ítem, y las características específicas de los grupos encontrados en estas variables, implementando procedimientos de análisis estadístico que respondan más cabalmente a las características del instrumento *portafolio*.

I INTRODUCCIÓN

Actualmente, los procedimientos orientados a garantizar la validez de los instrumentos de medición han incorporado, de forma más o menos generalizada, estudios de Funcionamiento Diferencial de Ítems (DIF, por su sigla en inglés), cuyo objetivo es determinar, mediante diferentes mecanismos de análisis estadístico, si las diferencias observadas en los resultados de una medición corresponden o no a diferencias reales entre grupos con ciertas características distintivas.

La aplicación de un análisis de Funcionamiento Diferencial de Ítem al portafolio se relaciona con una problemática que tiene gran relevancia en la actualidad, puesto que los resultados de este instrumento tienen una alta ponderación dentro del sistema de evaluación docente. Este, a su vez, aparece como un importante referente de la calidad de educación pública, por lo que constituye un insumo importante en la orientación de las políticas públicas para este sector. Considerando que la toma de decisiones en el ámbito de la educación tiene efecto, tanto en cada uno de los actores involucrados en sistema como en la sociedad en su conjunto y dadas las altas consecuencias asociadas a un sistema de evaluación aplicado a todos los profesores del sistema municipal, se requiere que dicho proceso de garantías de validez y equidad.

En cuanto a la organización del presente trabajo, es posible identificar cinco secciones. En primer lugar, el estudio se centrará en la revisión de los antecedentes relacionados con la aplicación de un sistema de evaluación de desempeño profesional docente en Chile, así como en las características del mismo. También, se revisarán el contexto actual de aplicación de la Evaluación, las posibles implicancias políticas que tienen en el actual escenario discusiones en torno a reformas legales en el campo educativo y el diseño de una política nacional docente. Al término de esta sección, se definirán los límites del problema de investigación y se declararán los objetivos que se

busca lograr a través de la misma.

La segunda sección corresponde a la revisión, en profundidad, de la literatura especializada referida a aspectos específicos del instrumento portafolio, además de la revisión de algunos resultados arrojados por el sistema de Evaluación Docente desde su implementación. En esta sección, también se profundizará en los procedimientos de análisis utilizados.

En la tercera sección, se explicará en detalle la metodología utilizada para el estudio, que consiste en la aplicación complementaria de dos procedimientos de análisis estadístico: Test Generalizado de Mantel-Haenszel, para ítems politómicos (QMH1 y QMH2) (Landis, Heyman, & Koch, 1978), y un índice obtenido a partir de la comparación de medias estandarizadas (SMD) (Zwick & Thayer, 1996; Zwick, Thayer, & Mazzeo, 1997).

La cuarta sección corresponde a la presentación de los resultados obtenidos a partir de los análisis en cada una de las variables en estudio y la discusión en torno a dichos resultados. Finalmente, en la quinta sección, se exponen las conclusiones extraídas a partir de la discusión de los resultados.

1.1 Antecedentes

El actual debate que se ha suscitado en torno a la educación pública, en este país, está lejos de tener un origen espontáneo o circunstancial; más bien, puede entenderse como parte de un proceso deliberativo que surge a mediados de la década de los 90 y que ha derivado en la implementación, a lo largo de las últimas dos décadas, de una serie de políticas públicas. En este sentido, ya el año 1996, en el discurso presidencial del 21 de mayo, se declaraba la voluntad del gobierno en ejercicio de promover la profesionalización de los docentes en el marco de una serie de reformas que consideraba iniciativas como el “Programa de Fortalecimiento de la Formación Inicial Docente, el de

Perfeccionamiento Fundamental, los programas de becas para estudiantes destacados que optaran por seguir la carrera de pedagogía, las pasantías y diplomados en el exterior y los premios de excelencia docente” (Bonifaz, 2011, p.17).

La Reforma Educacional, reimpulsada a fines de los años 90, vuelve a situar el tema de la calidad en el centro de las demandas sociales durante los primeros años del presente siglo. De acuerdo con lo señalado por Bellei, Contreras y Valenzuela (2008), ya en el año 2006, durante el primer período presidencial la presidenta Michelle Bachelet, fue convocado un amplio Consejo asesor por la Calidad de la Educación en respuesta a la gran agitación social provocada por las demandas en esta materia que habían sido enarboladas por el movimiento estudiantil. Este primer gesto institucional hacia el cambio ponía énfasis en la palabra *calidad*, como una deuda pendiente frente al avance en cobertura que había alcanzado el sistema durante la década anterior. En este escenario es que se reinstala en la agenda una de las aristas relacionadas con el déficit cualitativo: el que se relaciona con la calidad de los docentes y su relación con los aprendizajes de los estudiantes (Bellei et al., 2008).

Una serie de políticas orientadas al fortalecimiento de la profesión docente ha sido implementada a lo largo de las últimas décadas. Algunas de estas iniciativas se centraban en el fortalecimiento de la formación inicial, otras en el control y fortalecimiento del desempeño individual de los profesores y otras en promover el logro de objetivos institucionales a nivel de escuelas. Tal es el caso del Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educacionales Subvencionados (SNED), vigente desde 1996 mediante la ley n°19410, el cual estaba “destinado a incentivar la calidad de la gestión docente mediante un beneficio económico para los profesores de los establecimientos de mejor desempeño en cada región del país” (Rodríguez, 2015, p.12). En la lógica de los incentivos institucionales, también cabe señalar la Asignación de Desempeño Colectivo, consignada en el Artículo 18 de la Ley N° 19.933, modificada el año 2006. Esta disposición legal determinaba “un incentivo

monetario destinado a reconocer a los Equipos Directivos y Técnico-Pedagógicos, que avanzan en el mejoramiento de sus prácticas de gestión y liderazgo, generando más y mejores condiciones en la comunidad educativa, para lograr aprendizajes significativos y permanentes en el tiempo, en todos los estudiantes de sus establecimientos” (MINEDUC, 2014, p.6).

Otra línea de acción, es la implementación del programa INICIA, como una forma de fortalecer la calidad de la Formación Inicial Docente. Este programa tenía 3 componentes: definición de orientaciones curriculares, evaluación diagnóstica de conocimientos y competencias de los egresados de carreras de formación pedagógica y un programa de apoyo a las instituciones formadoras de profesores (García-Huidobro, 2008); sin embargo, en su implementación, se puso el énfasis en la evaluación de los egresados, pasando, de este modo, a ser la Prueba INICIA (García-Huidobro, 2008), cuyo objetivo también transitó desde *fortalecer* hacia *verificar* la calidad de la Formación Inicial Docente (MINEDUC, 2013). La aplicación de esta prueba se realizó entre los años 2008 y 2015, en aquellas instituciones de formación que optaban voluntariamente a su rendición. El año 2016 se suspende la aplicación de esta prueba para ser reemplazada por dos evaluaciones contempladas en un nuevo cuerpo legal, que crea el Sistema de Desarrollo Profesional Docente (Ley N° 20903, 2016). De acuerdo con esta nueva ley, las instituciones que imparten carreras de formación pedagógica deberán aplicar dos evaluaciones diagnósticas de carácter formativo, la primera al comienzo de la carrera, y la segunda durante los doce meses anteriores al último año de carrera, con el fin de orientar acciones de nivelación y acompañamiento que cada institución deberá implementar (Ley N° 20903, 2016).

En el marco de las iniciativas orientadas al reconocimiento del mérito individual de los profesores, se destaca el programa de Asignación de Excelencia Pedagógica (AEP). Esta iniciativa del Ministerio de Educación entró en vigencia el año 2002, como parte de una estrategia orientada al fortalecimiento de la profesión docente (Rodríguez,

2015). Este programa fue diseñado para evaluar dos grandes dimensiones del quehacer profesional docente: 1) conocimientos disciplinares y 2) habilidades pedagógicas generales de acuerdo con los estándares definidos en el Marco para la Buena Enseñanza. Para acceder a la certificación del programa AEP, los participaron voluntariamente de un proceso que consideraba dos instrumentos de evaluación: una prueba de conocimientos disciplinares y pedagógicos más un portafolio (Argüelles, Castillo, & Saragoni, 2015). La certificación de las competencias evaluadas por el programa implicaba para los docentes el derecho a percibir un beneficio económico de acuerdo con tres tramos, con una vigencia de 4 años. Complementariamente, los profesores acreditados accedían a la posibilidad de postular para formar parte de la Red Maestros de Maestros, una red de colaboración profesional creada para contribuir al desarrollo profesional de los docentes en ejercicio (Rodríguez, 2015). La continuidad del programa AEP se mantuvo vigente hasta el año 2015, año en que se suspendió su implementación, en el marco de la entrada en vigencia de la nueva política nacional docente, el año 2016.

A las iniciativas mencionadas anteriormente, se suma una de las políticas que marca un verdadero hito en las iniciativas de fortalecimiento docente y que ha significado un importante desafío en cuanto a los requerimientos técnicos necesarios para su implementación, así como en la demanda de los consensos necesarios que permitieran su puesta en práctica: el Sistema de Evaluación del Desempeño Profesional Docente (o Evaluación Docente) (Bonifaz, 2011). El sistema de educación municipal de Chile aplica un sistema de Evaluación del Desempeño profesional docente desde el año 2003, este sistema de evaluación aumentó progresivamente su cobertura hasta alcanzar a cubrir la totalidad de las escuelas municipales el año 2005.

La Evaluación Docente corresponde a un tipo de evaluación de desempeño basado en estándares, ello significó que, para su implementación, fuese necesaria la definición de esos estándares, los que fueron sistematizados en torno al Marco Para la Buena Enseñanza (MBE). El MBE es un documento en el que se definen y organizan

dominios y criterios, así como los descriptores más específicos en que estos se descomponen, estableciendo los principales aspectos por los cuales los docentes deben ser evaluados para identificar su nivel de desempeño (Bonifaz, 2011). El sistema de Evaluación Docente consiste en la aplicación de cuatro instrumentos: la confección de un portafolio, una entrevista realizada por un evaluador par, un informe de referencia de terceros y una autoevaluación. Cada uno de ellos tiene una ponderación específica definida legalmente en el Reglamento de la Evaluación Docente. Estos instrumentos se integran para determinar el nivel de desempeño final de los docentes (Flotts & Andrea, 2011). El portafolio es el instrumento cuya ponderación es más alta: pues su puntaje alcanza el 60% en el cálculo de la puntuación final. Este instrumento permite recoger directamente muestras del trabajo docente. Se estructura en dos módulos: uno en que se presenta información referida al diseño, implementación y evaluación de una unidad pedagógica; otro que consiste en la filmación de una clase (Flotts & Andrea, 2011). Las diferentes tareas del portafolio se organizan en función de 7 dimensiones, las que, a su vez, se subdividen en veinte indicadores.

Cabe señalar que la experiencia internacional contempla algunos ejemplos en los que se ha adoptado el uso de portafolios como parte de la medición de la calidad docente. En Singapur, el proceso de evaluación contempla una serie de instancias cuya evaluación final se basa en portafolios de trabajo (OCDE, 2009). Otro ejemplo de lo anterior es el caso de los EEUU, donde la evaluación PRAXIS, aplicada por el *Educational Testing Service* (ETS), incluye este tipo de instrumentos además de pruebas disciplinares y pedagógicas. En esta perspectiva, lo mismo ocurre con distintos sistemas de acreditación de estándares de práctica docente como *el Interstate New Teacher Assessment and Support Consortium* (InTASC), y con los estándares del *National Board for Professional Teaching Standards* (NBPTS) (Martínez-Rizo, 2012). También es posible mencionar el sistema de evaluación EdTPA, enfocado en estudiantes de pedagogía en práctica, y que ha sido adoptado por 25 estados en Estados Unidos (Martínez, 2013). Al mismo tiempo, en el ámbito nacional, existe evidencia que vincula los resultados obtenidos en el portafolio aplicado en el sistema de Evaluación Docente y

los aprendizajes de los estudiantes. Un reciente estudio que correlaciona los resultados en evaluación docente y mediciones de efectividad docente en 150 profesores de matemática y 162 profesores de lenguaje, concluyó que “los profesores más efectivos, es decir, aquellos que promueven mayores progresos en los aprendizajes de sus estudiantes, son quienes poseen mejores resultados en el portafolio, y por el contrario, los docentes menos efectivos obtienen más bajos resultados en este instrumento.” (Taut et al., 2014; Taut, Jiménez, & Barrientos, 2015 p.4). La estimación de la efectividad se obtuvo mediante una metodología de valor agregado, y considera, además de los resultados en la evaluación docente obtenida por los maestros el año 2006, los resultados de 5445 estudiantes, en el caso de los profesores de matemática, y 5998 estudiantes, en el caso de los profesores de lenguaje, en dos aplicaciones consecutivas de la prueba SIMCE (2004 y 2006) (Taut et al., 2015) .

Otro de los instrumentos que completa la evaluación y sus respectivas ponderaciones es la entrevista por un evaluador par (20%), que consiste en una serie de preguntas diseñadas a partir del MBE, aplicadas por un docente evaluador que debe desempeñarse en el mismo nivel, sector y modalidad de enseñanza del evaluado (Flotts & Andrea, 2011). El docente evaluador debe aplicar una rúbrica para asignar a cada respuesta uno de los cuatro niveles de desempeño definidos para la evaluación: Insatisfactorio, Básico, Competente o Destacado (Flotts & Andrea, 2011). Se suma al proceso el informe de referencia de terceros (10%): una pauta estructurada en base a una rúbrica, aplicada por un superior jerárquico (director o jefe de UTP) (Flotts & Andrea, 2011). Finalmente, se aplica una pauta de autoevaluación (10%). La ponderación cambia cuando se trata de la segunda o tercera evaluación de los docentes que han obtenido un resultado insatisfactorio: en esos casos, el portafolio se pondera en un 80%, la autoevaluación en un 5%, la entrevista en un 10% y los informes de referencia de terceros en un 5% (Flotts & Andrea, 2011).

Los resultados aportados en los casi 12 años de aplicación han servido como pauta para una serie de decisiones que involucran desde consecuencias individuales para

los docentes evaluados según sus niveles de logro hasta la elección de programas de apoyo y perfeccionamiento orientados a mejorar el desempeño profesional de los docentes con resultados insatisfactorios, por lo que constituye un proceso evaluativo con altas consecuencias a nivel individual y para el sistema. Las consecuencias asociadas a buen desempeño contemplaban, hasta el año 2015, la posibilidad de postular al beneficio económico denominado Asignación Variable por desempeño Individual (AVDI), cuyo monto correspondía a una proporción de la Renta Básica Mínima Nacional. Para acceder a este beneficio, el docente con desempeño Competente o Destacado en la Evaluación Docente, tenía una única posibilidad de rendir una prueba de conocimientos pedagógicos y disciplinarios y aprobarla con una calificación Suficiente, Competente o Destacado (Cortés & Lagos, 2011). El programa AVDI, al igual que AEP, fueron suprimidos con la entrada en vigencia de la ley N°20.903. Por otra parte, un bajo desempeño en la Evaluación Docente trae consecuencias que, en el caso de un resultado insatisfactorio reiterado, puede significar para el docente evaluado, la salida del sistema educativo.

Los profesores con un resultado Básico o Insatisfactorio son derivados a Programas de Superación Profesional (PSP); en el caso de aquellos con el nivel más bajo, deben reevaluarse el año siguiente. En este sentido, la Ley 20.501 de Calidad y Equidad en la Educación, de febrero de 2011, introdujo cuatro importantes modificaciones relacionadas con las consecuencias de la Evaluación Docente: reduce las oportunidades para superar un resultado Insatisfactorio de tres a dos evaluaciones consecutivas; en caso de no lograr superar este nivel de desempeño, el docente será desvinculado del sistema de educación municipal. Al mismo tiempo, la mencionada disposición legal disminuyó, para los docentes con resultado Básico, de cuatro a dos años el período entre evaluaciones, estableciendo la exigencia de superar dicho resultado en un máximo de tres evaluaciones consecutivas, debiendo en caso contrario hacer abandono de la dotación docente Por último, otorga a los directores de establecimientos la facultad de despedir anualmente hasta un 5% de la dotación docente entre aquellos profesores cuyo resultado vigente en la Evaluación Docente sea Básico o Insatisfactorio (Bonifaz, 2011; Cortés & Lagos, 2011).

Es importante señalar que los resultados reportados por la Evaluación Docente han presentado diferencias para determinados grupos de docentes; diferencias que diversos estudios han intentado explicar tanto a partir de características propias de los docentes como por aquellas que son contextuales (Bravo, Falck, González, Manzi, & Peirano, 2008; Valencia & Manzi, 2011). Considerando la amplia utilización de estos resultados, tanto en términos de consecuencias profesionales para los evaluados como en la orientación de políticas públicas, resulta fundamental determinar si las diferencias observadas derivan de que los atributos medidos efectivamente se distribuyen de forma heterogénea en diferentes grupos de docentes o, bien, son el resultado de una distorsión originada por el instrumento utilizado para dicha medición.

En el ámbito de las políticas públicas recientes destaca que, como parte de la nutrida agenda en educación que ha sido puesta en marcha por el segundo gobierno de la presidenta Michelle Bachelet, y con la intención de mejorar los resultados nacionales en esta materia frente al escenario internacional, se han comenzado a implementar una serie de modificaciones legales que abarcan reformas en diversos ámbitos del sistema escolar. En este sentido, la ley n° 20903 que crea un sistema de Desarrollo Profesional Docente incorpora una serie de modificaciones al marco jurídico que ha regulado el quehacer de los profesores hasta la fecha, además de regular las condiciones tanto de su formación inicial como de su formación en servicio.

Entre los aspectos más relevantes de este nuevo marco jurídico destaca la definición de un sistema progresivo de desarrollo profesional con cinco niveles que pueden ser recorridos por los docentes a lo largo de su trayectoria. La transición entre los diferentes niveles de este sistema se articula con el sistema de Evaluación Docente, considerando los resultados obtenidos en el portafolio de dicho proceso evaluativo, además de un nuevo instrumento diseñado para medir competencias pedagógicas y disciplinarias de los docentes. Cabe señalar que, como parte de la discusión de este proyecto se recogieron inquietudes planteadas por el Colegio de Profesores en cuanto a la necesidad de revisar la actual Evaluación Docente, e incorporar en el portafolio

elementos que midan aspectos relativos a trabajo colaborativo y funciones profesionales no contempladas (BCN, 2016). En este sentido, la evaluación docente del año 2016 incorporó ciertas modificaciones en el portafolio, agregándose un tercer módulo que mide las dimensiones de *Trabajo colaborativo*, *Actividades de perfeccionamiento* y *Responsabilidades profesionales*. La aplicación de este instrumento, con las modificaciones mencionadas, se encuentra actualmente en proceso, esperándose los primeros resultados para marzo de 2017, razón por la cual no es posible considerar dichas modificaciones en el presente estudio.

Otros aspectos relevantes de esta ley son: la definición requisitos de exigencia para acceder a estudiar carreras de pedagogía, al mismo tiempo, se determina que las instituciones formadoras deben contar con una acreditación basada en criterios de calidad, a cargo de la Comisión nacional de Acreditación. En cuanto a la formación en servicio, se consagra el derecho a acceder a programas de perfeccionamiento gratuito para todos los docentes del sistema, de acuerdo con sus necesidades y las de su comunidad educativa. Se modifican aspectos relacionados con las condiciones laborales, aumentando la cantidad de horas no lectivas, e incrementando las remuneraciones de los profesores; además de lo anterior, se establece un sistema de inducción y acompañamiento para docentes principiantes (Ley N° 20903, 2016).

Al observar la nuevas normas que se han implementado en educación, se observa que un aspecto fundamental está justificado por la existencia de un sistema de evaluación que aporte información para vincular el progreso de los docentes dentro del sistema con aspectos del desarrollo profesional que aporten calidad al sistema. En este perspectiva, la aplicación de mediciones estandarizadas ha pasado a formar parte de las herramientas indispensables para la obtención de información acerca de habilidades, actitudes, desempeños y opiniones de las personas y grupos de personas, información requerida en diversas disciplinas, entre ellas, por supuesto, la educación (Gómez-Benito, Hidalgo, & Guilera, 2010). En este sentido, y siendo la Evaluación Docente una medición cuya aplicación ha adquirido gran protagonismo en las actuales discusiones en

torno a la calidad de la educación, cabe señalar que Chile no es un caso aislado en cuanto a la aplicación de instrumentos de medida estandarizados en los procesos de evaluación, así como en los procesos de recogida de información para orientar la toma de decisiones a nivel de políticas públicas e institucionales. En la experiencia internacional, la aplicación de instrumentos como portafolios y pautas de observación de clases es una herramienta ampliamente extendida (Martínez-Rizo, 2012; Martínez, 2013).

Frente a lo anterior, hay tres aspectos insoslayables sobre este tema. En primer lugar, dada la importancia de las decisiones basadas en el sistema de Evaluación Docente, este debe contar con los más rigurosos estándares que permitan la mayor validez posible en sus resultados garantizando, así, la justicia de las decisiones tomadas a partir de los mismos. Por otro lado, el proceso de validación es permanente y genera evidencia acumulativa que contribuye al constante mejoramiento y adecuación de los instrumentos evaluativos; en este sentido, el presente estudio pretende aportar con una nota en el desarrollo polifónico de las investigaciones en torno a la Evaluación Docente. Por último, de cara al escenario educativo que se comienza a configurar, la vasta acumulación de evidencia respecto de la validez de la Evaluación Docente (Taut et al., 2015, 2012) impone al sistema el desafío de articular los instrumentos evaluativos que lo conforman con los nuevos requerimientos que el nuevo contexto exija, lo que demandará la revisión de los instrumentos actuales a la luz de la evidencia acumulada y las nuevas necesidades.

1.2 Delimitación del problema de investigación

Considerando el importante papel que juega la Evaluación Docente en la toma de decisiones, además de sus consecuencias que involucra a nivel individual, resulta de gran relevancia que la información sea recogida por instrumentos que cumplan con los criterios de rigor necesarios para garantizar un proceso válido, objetivo y confiable. De

acuerdo con Taut, Santelices, & Manzi (2011), solamente se han reportado análisis preliminares de Funcionamiento Diferencial de ítems a partir de los resultados del Portafolio 2009 mediante el método de Mantel-Haenszel y análisis de multivarianza. De acuerdo con el estudio mencionado, no se encontró evidencia de funcionamiento diferencial entre docentes de establecimientos rurales y urbanos, como tampoco entre hombres y mujeres; sin embargo, el mismo artículo señala que “en el futuro se profundizará el estudio de sesgo de ítem en la Evaluación Docente” (Taut et al., 2011, p.166). Cabe señalar que, fuera del artículo antes mencionado, no se han encontrado reportes acerca de otros análisis de DIF, ya sea para profundizar el análisis sobre las variables ya analizadas, o bien, que consideren aquellas variables que han reportado diferencias significativas en los resultados del portafolio; tales como la edad, la asignatura impartida por el docente evaluado o el nivel de enseñanza en el cual éste se desempeña. En esta misma línea, sería de gran utilidad, en términos de la especificidad de la información aportada, que el análisis se realice a nivel de cada uno de los veinte indicadores del Portafolio, los que, a su vez, se encuentran agrupados en 7 dimensiones. En este sentido, resulta especialmente pertinente plantearse las siguientes interrogantes: ¿Son estadísticamente significativas las diferencias de desempeño que se presentan al analizar los resultados en función de variables como sexo, edad, zona geográfica, nivel de enseñanza y asignatura que imparte el docente? ¿Son esas diferencias la consecuencia de una distorsión producida por las características psicométricas del Portafolio o, bien, responden a una distribución heterogénea de los atributos medidos entre los miembros de la población? Siendo este el instrumento con mayor ponderación dentro del sistema de Evaluación Docente, además de ser el único que aporta información directa del quehacer profesional docente (Sun, Correa, Zapata, & Carrasco, 2011), despejar aquellas dudas pasa a ser parte de un necesario proceso de evaluación a la Evaluación Docente, lo que enmarca el presente trabajo entre los estudios de validez de este sistema evaluativo.

1.3. Objetivos

1.3.1 Objetivo General

Analizar la presencia o ausencia de Funcionamiento Diferencial de los Ítems (DIF) en la medición de los indicadores evaluados en el portafolio de la Evaluación del Desempeño Profesional Docente 2013, de acuerdo con las variables *Sexo*, *Edad*, *Zona Geográfica*, *Nivel de Enseñanza* y *Asignatura*.

1.3.2 Objetivos Específicos.

1. Determinar la presencia de diferencias, estadísticamente significativas, en la medición de los indicadores del portafolio para las variables *Sexo*, *Edad*, *Zona Geográfica*, *Nivel de Enseñanza* y *Asignatura* que imparte el docente.
2. Determinar la presencia de DIF en los indicadores del portafolio de la Evaluación del Desempeño Profesional Docente, para la variable *Sexo*.
3. Determinar la presencia de DIF en los indicadores del portafolio de la Evaluación del Desempeño Profesional Docente, para la variable *Zona Geográfica* (rural-urbano).
4. Determinar la presencia de DIF en los indicadores del portafolio de la Evaluación del Desempeño Profesional Docente, para la variable *Edad*.
5. Determinar la presencia de DIF en los indicadores del portafolio de la Evaluación del Desempeño Profesional Docente, para la variable *Nivel de Enseñanza*.
6. Determinar la presencia de DIF en los indicadores del portafolio de la Evaluación del Desempeño Profesional Docente, para la variable *Asignatura* que imparte.

II MARCO DE REFERENCIA

El presente capítulo se organiza en cuatro grandes secciones, presentándose, en primer lugar, una revisión en torno a diferentes formas en que se mide la calidad docente, entendiendo que dichas formas de medición responden, a su vez, a diferentes enfoques en la discusión sobre la calidad docente, y sus respectivas conceptualizaciones y utilizadas para operacionalizar este constructo.

A continuación, se presenta una revisión acerca de las características del sistema de evaluación de desempeño profesional docente aplicado en Chile, profundizando específicamente en las características del instrumento portafolio, sobre cuyos resultados se realizarán los análisis de este estudio.

Una tercera sección de este capítulo se centra en revisar la literatura en torno a los aspectos teóricos que sustentan esta investigación. En primer lugar, se aborda la conceptualización del Funcionamiento Diferencial del Ítem (DIF), así como las posibles clasificaciones del mismo y los métodos y técnicas utilizadas para su detección.

Finalmente, se presenta una revisión de resultados reportados por la Evaluación Docente desde su implementación a la fecha, en función de variables que han presentado diferencias sistemáticas que justifican un estudio de funcionamiento diferencial sobre este instrumento.

2.1 Medición de la Calidad Docente

La necesidad de medir el desempeño docente surge de la evidencia que han aportado numerosos estudios que sostienen que el protagonismo de los docentes como principales mediadores en el proceso de interacción entre los estudiantes y el conocimiento (Avalos, 2002; Avalos, Cavada, Pardo, & Sotomayor, 2010). Esta

evidencia señala que la calidad de la enseñanza que la escuela imparte, así como los resultados de aprendizaje de los estudiantes, se relacionan entre otras variables con las características personales y profesionales de los profesores, las que, a su vez se vinculan con la forma en que enfrenta y conduce el proceso pedagógico (Fang, 1996; Vera, Osses, & Schiefelbein, 2012).

Sin embargo, la discusión en torno a la calidad de la educación no es concluyente con respecto a los factores asociados a la variabilidad de los resultados escolares y al porcentaje de dicha variabilidad que puede atribuirse a diferencias entre los docentes que atienden a los estudiantes (Lara, Mizala, & Repetto, 2010). En este mismo sentido, es necesario considerar que las múltiples diferencias en la definición de los constructos de calidad y competencia docente dependen, en gran medida, “del uso que se le quiera dar y del contexto específico en que se sitúe este uso” (Martínez, 2013, p.8). Por otra parte, pese a la dificultad para definir la calidad docente, puesto que esta se encontraría modulada por el contexto (Santelices, Galleguillos, & González, 2015), existe un amplio consenso sobre la importancia de la calidad de enseñanza en el aprendizaje de los estudiantes (Barber & Mourshed, 2008; Santelices et al., 2015; Wilson, 2009).

De acuerdo con lo señalado por Santelices et al. (2015), existen tres líneas de investigación en la literatura internacional que abordan la discusión sobre calidad docente: el estudio de características observables del profesor, el análisis de prácticas profesionales basadas en estándares profesionales y estimaciones de valor agregado, basadas en el rendimiento de los estudiantes en pruebas estandarizadas (Santelices et al., 2015).

Otras propuestas han incorporado recientemente sistemas evaluativos que integran diferentes instrumentos y fuentes de información (Martínez, 2013) Este enfoque considera que la calidad docente constituye un constructo más complejo, donde las diferentes aproximaciones corresponden a medidas de las subdimensiones del mismo (Martínez, 2013; Santelices et al., 2015).

Las mediciones de valor agregado y el análisis de las prácticas docentes han aportado una amplia evidencia que permite relacionar determinadas prácticas docentes con un impacto en el resultado reportado por las mediciones aplicadas al aprendizaje de los estudiantes. Por ejemplo, un estudio realizado en Chile el año 2010 revela que leer textos, realizar ejercicios de comprensión lectora y llevar a los niños a la biblioteca se correlacionan con un mejor desempeño en el SIMCE de Lenguaje. El mismo estudio señala que los docentes que declaran altas expectativas en sus estudiantes obtienen en sus puntajes adicionalmente un 12% de una desviación estándar en las puntuaciones SIMCE de Lenguaje y Matemática (Lara et al., 2010). Diversos trabajos han intentado establecer una relación entre características de los docentes y resultados académicos de los estudiantes en el contexto nacional (Bravo et al., 2008; Lara et al., 2010; León, 2008; Toledo, 2010). Esta tendencia no es otra cosa que la expresión de una necesidad por identificar aquellas características de los docentes cuyo desarrollo permitiría garantizar un incremento en los indicadores de calidad aplicados sobre el sistema educativo. Del mismo modo, identificar aquellas características permitiría implementar programas orientados al desarrollo de habilidades específicas en los docentes, ya sea como parte de su formación inicial y/o continua.

A nivel internacional, la evidencia obtenida con metodologías de valor agregado asigna a las características de los docentes una gran importancia en el intento por explicar las diferencias en los aprendizajes de los estudiantes. En el estudio *The Market for Teacher Quality* (Hanushek, Kain, O'Brien, & Rivkin, 2005), se analizaron datos para tres cohortes de estudiantes del estado de Texas, donde se encontraron diferencias significativas entre colegios en términos del aprendizaje de sus alumnos explicadas fundamentalmente por el efecto de los profesores más que por diferencias en la organización de los colegios, en liderazgo o en sus condiciones financieras. En particular, sugieren que las diferencias en la calidad de los profesores explican al menos un 7.5% de la variación total en los resultados de los alumnos (Hanushek et al., 2005; Lara et al., 2010).

Los modelos de valor agregado aplicados a la medición de la calidad docente han tenido una serie de cuestionamientos tanto por la dificultad de su implementación y las consecuencias que esto puede generar en el sistema. La asignación no aleatoria de estudiantes a profesores supone un problema para este tipo de estudios ya que puede generar distorsiones en la estimación de los efectos de la escuela y profesores (Baker et al., 2010; H. Braun, Chudowsky, & Koenig, 2010; Steele, Hamilton, & Stecher, 2010). Por otra parte se observa la necesidad de contar con mediciones longitudinales y la imposibilidad práctica de aislar el efecto de otras variables en el desempeño escolar. Del mismo modo, los modelos de regresión sobre los que se basan los análisis de valor agregado requieren satisfacer una serie de supuestos que con frecuencia no se cumplen en la realidad educativa, un ejemplo de esto es que las pruebas se reportan en una escala de intervalo, aun cuando esto no se verifique estrictamente en la mayoría de los casos (Braun et al., 2010). Para una información más completa en cuanto a las características de estos modelos y sus potenciales limitaciones, consultar Taut et al. (2012).

A pesar de la existencia de múltiples estudios que buscan aproximarse a una medición de la calidad docente a través de los resultados de los estudiantes, a veces utilizando metodologías de valor agregado, otras con procedimientos multivariantes y multinivel, que buscan establecer correlaciones entre características específicas de los docentes y el desempeño de sus estudiantes, el sistema escolar chileno adoptó un modelo de evaluación de desempeño profesional docente basado en un cuerpo de estándares de desempeño, descritos en el Marco Para la Buena Enseñanza.

La elaboración del Marco para la Buena Enseñanza (MBE) tomó como base los estándares de desempeño para la formación docente, cuyos lineamientos fueron ampliados y contextualizados a la realidad de los docentes en servicio (Avalos, 2002; Bonifaz, 2011). Para dar cuenta del conjunto de responsabilidades que los docentes enfrentan en su trabajo diario, la definición de los estándares del MBE se articuló en torno a ciertas preguntas tales como “¿Qué es necesario saber?”, “¿Qué es necesario saber hacer?” y “¿Cuán bien se debe hacer?” o “¿Cuán bien se está haciendo?”. Las

respuestas a estas preguntas se tradujeron en el desarrollo de un marco conformado por cuatro dominios que agrupaban a su vez veinte criterios de ejercicio profesional, cuya función principal era orientar el trabajo profesional del docente de aula. Los dominios y criterios del MBE, así como sus descriptores, establecen los principales aspectos por los cuales los docentes deben ser evaluados, permitiendo, a su vez, determinar su nivel de desempeño. El MBE fue sometido a una consulta pública por establecimiento, en la que participaron 73.000 docentes agrupados en 6.696 grupos de trabajo. Esta consulta se desarrolló en dos etapas: la primera de ellas fue realizada en diciembre de 2001, con la participación de 2.371 grupos de profesores; la segunda, en abril de 2002, y en ella participaron otros 4.325 grupos (Bonifaz, 2011).

Los Dominios del MBE y sus respectivos criterios se grafican en la Figura n°1

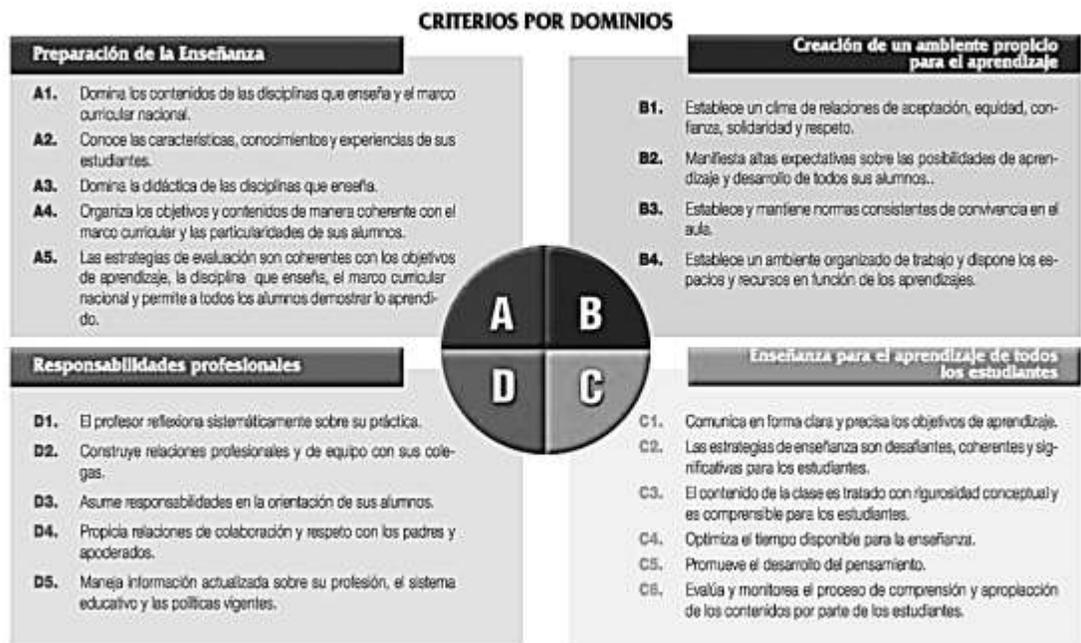


Figura n°1. Organización de los criterios del MBE. (MINEDUC, 2008)

2.2 Evaluación del Desempeño Profesional Docente en Chile

Desde el año 2003, se comenzó a implementar en Chile un programa nacional de evaluación de profesores, el Sistema de Evaluación del Desempeño Profesional Docente, para alcanzar cobertura nacional el año 2005, abarcando a todas las escuelas de administración municipal, con el fin de incorporar, de este modo, a profesores de todas las comunas del país. Más de 67.000 docentes que ejercen en los principales niveles, modalidades y subsectores de enseñanza han participado en el proceso y alrededor de 14.000 de ellos ya habían completado su segunda evaluación hacia el año 2011 (Sun, Calderón, Valerio, & Torres, 2011). Asimismo, existe amplia coincidencia en que el principal impulsor de las variaciones en el aprendizaje escolar es la calidad de los docentes (Barber & Mourshed, 2008). Por ello, resulta fundamental contar con un sistema válido y confiable para evaluar su desempeño; asimismo, promover una cultura de evaluativa que promueva la reflexión y mejoramiento continuo, a partir de la interacción profesional entre docentes, y permita orientar la implementación de iniciativas de formación en servicio que impacten positivamente los aprendizajes de los estudiantes mejorando, así, la calidad de la educación chilena (Sun et al., 2011).

Este programa de evaluación del desempeño está basado, como ya se mencionó, en estándares, lo que implica la existencia de un modelo ideal de comparación, a partir del cual se definen los criterios que permiten a los docentes conocer las competencias que caracterizan un buen desempeño en aula en relación a las cuales serían evaluados (Bonifaz, 2011). Este cuerpo de estándares estuvo a cargo de un subcomité sobre estándares creado a partir de la comisión tripartita sobre evaluación docente.

El sistema de Evaluación del Desempeño Profesional Docente considera diferentes instrumentos, a partir de los que cada docente es evaluado asignándosele uno de los 4 niveles siguientes: *Insatisfactorio*, *Básico*, *Competente* y *Destacado*. Estos instrumentos se enuncian a continuación con sus respectivas ponderaciones: portafolio (evidencia escrita y clase filmada), 60%; entrevista del evaluador par, 20%, que consiste

en la administración de una pauta de entrevista estructurada, sobre la cual un evaluador par aplica una rúbrica para asignar uno de los cuatro niveles desempeño a cada respuesta del docente evaluado; pauta de autoevaluación, 10%; informe(s) de referencia de terceros, 10%, que consiste en una pauta estructurada que debe contestada por el director del establecimiento y por el jefe de UTP, quienes, por separado, evalúan al docente asignando un nivel de desempeño por cada pregunta de la pauta (Flotts & Andrea, 2011). De esta batería de instrumentos evaluativos, el portafolio es el instrumento que entrega los resultados más bajos, ya que desde 2007 a la fecha, menos del 20% de los docentes alcanzan niveles Competente o Destacado en este ámbito (Flotts & Andrea, 2011).

Cabe señalar que existen estudios que aportan una serie de cuestionamientos hacia algunos de los diferentes instrumentos que componen esta evaluación, por ejemplo, se ha señalado la existencia de sesgos asociados a la autoevaluación, lo que se refleja en el alto porcentaje de docentes que alcanza un desempeño destacado en este instrumento (Santiago, Benavides, Danielson, Goe, & Nusche, 2013; Taut et al., 2011). Lo mismo sucede con relación a la entrevista de evaluador par, donde se ha observado que la percepción subjetiva de los evaluadores tiende a favorecer a los docentes (Taut et al., 2011). Por otra parte, en cuanto a los informes de referencia de terceros, instrumento elaborado por superiores jerárquicos al evaluado, existe evidencia contradictoria. En el ámbito internacional que ha cuestionado el poder de los directivos para discriminar niveles de desempeño docente, señalando que los directores solamente tienen la capacidad (o preparación) para identificar aquellos profesores que se encuentran en los extremos de la distribución por rendimiento, es decir, pueden identificar a los buenos docentes y a los docentes con desempeños insatisfactorios. Mientras, en el contexto nacional, se señala que “de los cuatro instrumentos que integran el sistema de evaluación, el informe del Superior jerárquico es uno de los que mejor identifica los desempeños de los docentes” (MINEDUC, 2013).

2.2.1 Descripción del portafolio

Es importante señalar que, de los cuatro instrumentos contemplados en el sistema de Evaluación Docente, el portafolio es el único que aporta información directa del trabajo docente (Sun, Correa, et al., 2011). Está compuesto por dos secciones (módulos), el primero de ellos (Módulo 1) consiste en un reporte con evidencias de la implementación de una unidad de aprendizaje, en un nivel y asignatura en que el docente se desempeña; el profesor debe desarrollar un objetivo de aprendizaje contemplado en el currículum nacional, el cual es escogido a partir de dos objetivos de aprendizaje previamente determinados para cada asignatura y nivel de enseñanza. El material entregado por el docente corresponde a planificaciones, material utilizado para el desarrollo de las clases, evaluaciones de aprendizaje y la respuesta a una serie de preguntas relacionadas con la utilización del material, el análisis y utilización de los resultados evaluativos, así como la reflexión que el propio docente realiza en torno a su práctica docente (Araya, Taut, Santelices, & Manzi, 2011; Taut & Sun, 2014).

La segunda sección del portafolio (Módulo 2) consiste en el registro en video de una clase de 40 minutos realizada por el docente. La grabación del video se realiza por un camarógrafo externo, en una fecha y hora previamente informada al docente. Este módulo entrega información acerca del clima dentro de la sala de clases, la estructura de la misma, así como de las interacciones pedagógicas entre el docente y los estudiantes (Taut & Sun, 2014).

Hasta el año 2011 las tareas del portafolio se organizaban en torno a ocho dimensiones: Dimensión A: Organización de los elementos de la unidad; Dimensión B: Análisis de las actividades de las clases; Dimensión C: Calidad de la evaluación de la unidad; Dimensión D: Reflexión a partir de los resultados de la evaluación; Dimensión E: Reflexión pedagógica; Dimensión F: Ambiente de la clase; Dimensión G: Estructura de la clase y Dimensión H: Interacción pedagógica (Manzi, Meckes, Ortega, & Lacerna, 2011). Las cinco primeras correspondían al Módulo 1 y las tres restantes al Módulo 2.

Cada dimensión estaba operacionalizada en 3 indicadores, exceptuando las dimensiones D y H, con dos y cuatro indicadores respectivamente. A partir del año 2012, los indicadores correspondientes a las cinco dimensiones correspondientes al Módulo 1 fueron reorganizados en las 4 dimensiones que se mantienen vigentes hasta la actualidad (Taut & Sun, 2014) (Figura n°2.)

	TAREA	DIMENSIÓN	INDICADORES
MÓDULO 1	Descripción de una unidad pedagógica	1) Organización de los elementos de la Unidad	1.1) Formulación y coherencia entre los objetivos 1.2) Relación entre actividades y objetivos 1.3) Calidad de la secuencia entre las clases de la unidad
	Reflexión a partir de una unidad pedagógica	2) Análisis de las actividades de las clases	2.1) Características de los alumnos consideradas en la unidad
	Aprendizaje profesional a partir de una clase		2.2) Análisis de la unidad realizada 2.3) Análisis de una clase
	Evaluación de la unidad pedagógica	3) Calidad de la Evaluación de la Unidad	3.1) Calidad de la evaluación 3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada
	Reflexión a partir de los resultados de la evaluación Retroalimentación a partir de la evaluación	4) Reflexión a partir de los resultados de la Evaluación	4.1) Atribución de los resultados de aprendizaje 4.2) Calidad de la Retroalimentación
MÓDULO 2	Clase grabada	5) Ambiente de la clase para el aprendizaje	5.1) Promoción de la focalización de los alumnos en las actividades 5.2) Promoción de la participación de todos los alumnos 5.3) Supervisión de las actividades
		6) Estructura de la clase	6.1) Calidad del inicio de la clase 6.2) Calidad del cierre de la clase 6.3) Contribución de las actividades al logro de los objetivos de la clase
		7) Interacción pedagógica	7.1) Calidad de las explicaciones 7.2) Calidad de las preguntas 7.3) Calidad de la retroalimentación a los alumnos 7.4) Indicador por asignatura o área

Figura 2. Organización del portafolio. Fuente: www.docentemas.cl

El plazo para desarrollar el portafolio es de 12 semanas, a partir de entonces comienza el proceso de corrección, que se lleva a cabo por docentes de aula con un mínimo de 5 años de experiencia, y que se desempeñan en el mismo nivel y asignatura correspondiente al módulo que les corresponde revisar. Los correctores son sometidos a un proceso de capacitación y entrenamiento en la aplicación de las rúbricas y pautas de corrección, además, no tienen acceso a los datos de identificación personal del autor del portafolio, por lo que la revisión es anónima; finalmente, un 20% de los portafolios son seleccionados aleatoriamente para ser sometidos a una doble corrección, con el fin de garantizar la confiabilidad y objetividad del proceso.

A cada uno de los veinte indicadores contenidos en el portafolio se le aplica una rúbrica que clasifica el nivel de logro obtenido por el docente en una de cuatro categorías: *insatisfactorio*, *básico*, *competente* y *destacado*. Las dimensiones del instrumento se ponderan igualmente para determinar el puntaje final que se expresa en una escala de 1 a 4.

En los estudios reportados hasta el año 2011 se consignan diferencias entre la estructura factorial arrojada por el portafolio y las dimensiones utilizadas para reportar los resultados, como también con respecto a los dominios de MBE. Para el primer módulo se han identificado tres tipos de habilidades: “1) habilidad para diseñar y planificar la enseñanza, a veces dividida en (1a) planificación de la clase y (1b) diseño de actividades pedagógicas; (2) habilidad para construir instrumentos de evaluación; y (3) capacidad reflexiva sobre la propia práctica.” (Taut et al., 2011 p.163). Por otra parte, el segundo módulo ha mostrado consistentemente agruparse en dos factores altamente correlacionados, con excepción del año 2006, cuando el análisis arrojó 3 factores (Taut et al., 2011).

Taut et al. (2011) también reportan otros estudios que se han llevado a cabo sobre el portafolio analizando la confiabilidad entre jueces, para examinar el desempeño de los

correctores como una potencial fuente de error, y la generalizabilidad, para determinar la proporción de variabilidad de los puntajes que se atribuye a (1) diferencias “reales”; (2) diferencias sistemáticas atribuibles a los correctores, y (3) error “residual”. El estudio de confiabilidad entre jueces (IRR, por su sigla en inglés), realizado el año 2005 arrojó un índice de 0,061 en promedio para ambos módulos. Los estudios de generalizabilidad, fueron realizados los años 2008, 2009 y 2010, y en estos tres años, la varianza atribuible a diferencias reales fluctúa entre 25% y 50%, mientras que la varianza atribuida a los correctores corresponde entre un 3% y un 10%, variando según la asignatura y la dimensión del portafolio (Taut et al., 2011). El último estudio de generalizabilidad analizó la faceta del corrector en dos modalidades distintas (por profesor o por dimensión), mostrando que un 31% de la varianza se puede atribuir al efecto de los correctores, porcentaje que aumenta si se analiza por dimensión. Estos resultados indican que “la generalizabilidad es adecuada en lo que respecta al ordenamiento de los puntajes finales (decisiones relativas), pero que el actuar de los correctores no alcanza niveles óptimos para el caso en que las decisiones deben basarse en el puntaje observado por sí mismo (decisiones absolutas)”(Taut et al., 2011 p.165).

Pese a que la Evaluación Docente se ha insertado en la cultura de las escuelas municipales, no está exenta de cuestionamientos y dificultades. Una de las dificultades que enfrenta el proceso es la necesidad implementar una adecuada capacitación y entrenamiento para los correctores, y la longitud de tiempo que todo el proceso de revisión requiere; también se observa como una dificultad, la necesidad de comparar materiales y productos aportados por los docentes que fueron desarrollados con propósitos distintos (García, 2014). A lo anterior, se suma la generalizada percepción entre los docentes de sobrecarga laboral asociada a la falta de tiempo para desarrollar el portafolio (García, 2014; Taut et al., 2011; Taut & Sun, 2014).

2.3 Funcionamiento Diferencial de Ítems

En términos generales, un determinado ítem presenta DIF si se comporta diferencialmente para individuos o grupos comparables que difieren con respecto a alguna variable que pueda constituir una fuente sistemática de variación ajena al atributo medido por el ítem, entendiendo por “comparables” aquellos grupos de sujetos que poseen el mismo nivel en la habilidad o atributo medido por el instrumento (Gómez-Benito & Hidalgo, 1997, 2005; Gómez-Benito, Hidalgo, & Ferré, 2005).

El análisis para determinar la presencia de DIF se basa en la comparación entre el funcionamiento de los ítems para un grupo, denominado *focal*, que agrupa al conjunto de individuos sobre el que generalmente se centra el interés del análisis, y el funcionamiento de los ítems para el grupo de referencia, constituido por un grupo de sujetos estándar respecto al cual se compara el grupo focal (Gómez-Benito & Hidalgo, 1997; Holland & Thayer, 1988). En la mayor parte de los casos, se considera para el análisis un único grupo focal. Sin embargo, cabe la posibilidad de comparar el funcionamiento de los ítem para múltiples grupos focales respecto de un mismo grupo de referencia (Gómez-Benito et al., 2005).

La aplicación de análisis DIF a los diferentes instrumentos de medición ya estandarizados, como en los procesos de evaluación de nuevos instrumentos, radica en la necesidad de asegurar que la información recogida con dichos instrumentos cuenta con la validez que permita tomar decisiones sobre la base de que las mediciones no están favoreciendo o desfavoreciendo a alguno de los grupos estudiados. Esto concuerda con lo señalado por Gómez-Benito et al. (2005) al referir las recomendaciones publicadas en *Standards for Educational and Psychological Testing* (AERA, APA, 1999) en cuanto que, para todo instrumento de medición, los estudios para la detección del DIF deberían considerarse como una etapa propia del proceso de evaluación de su validez, recomendación consistente con lo señalado en los *ETS Standards for Quality and Fairness* (ETS, 2015).

Tradicionalmente, los análisis estadísticos que buscan identificar la presencia o ausencia de DIF se enmarcan dentro los estudios de validez y se utilizan para identificar diferencias en el funcionamiento de los ítems que pudieran ser indicios de sesgos originados en alguna característica del ítem de la prueba al producirse un error sistemático que pudiera distorsionar la medición de las habilidades o atributos. Sin embargo, es necesario señalar que DIF y *sesgo* no son palabras intercambiables, por lo que no deben ser utilizadas como sinónimos, sino más bien como conceptos complementarios.

La presencia de DIF puede tener diversos motivos, y determinar dichas causas ha sido el foco de múltiples investigaciones, centrande, por una parte, el estudio en aspectos propios de la composición y/o exposición diferencial de los subgrupos, y la consecuente interpretación del contenido del ítem (por ejemplo, contenido culturalmente específico, o lenguaje con un nivel de abstracción, o carga semántica particular); como también se ha examinado la relación entre los estadísticos del DIF y las características de los ítems y su contexto en el test, tales como, su posición relativa en la prueba o subtest, el tipo de ítem, entre otros (Carlton & Harris, 1992; Holland & Wainer, 1993; Longford, 2014; Santelices & Wilson, 2010; Schmitt & Bleistein, 1987). Actualmente, la literatura tiende a concordar en que, para determinar si un ítem presenta algún tipo de sesgo, el estudio deberá contemplar una primera etapa, consistente en determinar por medio de procedimientos estadísticos la presencia o ausencia de ítems con DIF, y otra etapa en que el análisis trasciende el ámbito de lo cuantitativo para intentar explicar las causas del funcionamiento diferencial (Camilli & Shepard, 1994; Herrera, 2005). Es decir, la sola presencia de DIF en un determinado ítem no implica necesariamente que dicho ítem esté sesgado y, por otra parte, es posible que un ítem con sesgo no necesariamente presente DIF (Camilli & Shepard, 1994). En este punto, es importante señalar que el presente estudio se acotará a la aplicación de procedimientos de análisis estadístico para la detección de ítems con DIF sin internarse en el análisis de los posibles factores que originen el funcionamiento diferencial.

Las diferencias en el desempeño de un test no significan que necesariamente se esté en presencia de funcionamiento diferencial, ya que las diferencias que presentan los resultados pueden ser la consecuencia que un grupo presenta efectivamente mayor habilidad o atributo que otro grupo, en este caso, la diferencia entre los grupos se denomina Impacto (Ackerman, 1992; Camilli & Shepard, 1994).

2.3.1 Tipos de DIF

Se distinguen dos tipos de DIF, determinados por la existencia o ausencia de interacción entre el nivel de habilidad o atributo medido y el grupo al que pertenecen los sujetos. Cuando no existe interacción entre el nivel de habilidad o atributo medido y la pertenencia a un grupo, es decir que la probabilidad de responder correctamente a un determinado ítem es mayor para todos los integrantes de un grupo independiente de su nivel de habilidad, se está en presencia de DIF uniforme. Por otra parte, cuando la probabilidad de responder correctamente un ítem para los integrantes de un grupo varía dependiendo del nivel de habilidad o atributo medido, se está en presencia de DIF no uniforme, en este caso existe la posibilidad de que el DIF a favor de un grupo, en un nivel de habilidad sea compensado por el DIF contra el mismo grupo, en otro nivel de habilidad (De Ayala, Kim, Stapleton, & Dayton, 1999; Gómez-Benito et al., 2005; Hidalgo & Gómez-Benito, 2000; Mellenbergh, 1982).

2.3.2 Métodos de detección del DIF

Son numerosas las técnicas que han sido desarrolladas para detectar el DIF. Siguiendo la taxonomía de estadísticos DIF propuesta por Potenza y Dorans (1995), se considerarán dos criterios principales para caracterizar los diferentes procedimientos disponibles a la hora de evaluar si los ítems de un test presentan funcionamiento diferencial: la forma de estimar el rasgo medido y la relación entre la puntuación en el

ítem y la variable de equiparación. El primer criterio considera dos posibilidades para la estimación del rasgo o habilidad sometido a medición; por una parte, encontramos los métodos que utilizan la puntuación observada en el test para definir procedimientos de equiparación entre los grupos y por otro, los procedimientos que consideran la relación entre los patrones de respuesta y una variable latente. Tal como señala Recio (2012), en el primer grupo de técnicas encontramos a toda la familia de estadísticos Mantel y Mantel-Haenszel (Holland & Thayer, 1988), estandarización (Dorans & Kulick, 1986), modelos de regresión logística (Swaminathan & Rogers, 1990), modelos loglineales (Mellenbergh, 1982), análisis discriminante logístico (Miller & Spray, 1993), método delta-plot (Angoff & Ford, 1973). El otro grupo reúne los procedimientos que utilizan la estimación de un rasgo latente, que es una función de los datos observados (Potenza & Dorans, 1995). De acuerdo con Recio (2012), en este grupo encontramos el estadístico de Lord (Lord, 1980), medidas de área (Raju, 1988, 1990) y SIBTEST (Shealy & Stout, 1993).

El segundo criterio para caracterizar los procedimientos de detección del DIF se define por la forma en que la puntuación del ítem se relaciona con la variable de equiparación de los grupos. Esta relación puede ser paramétrica es decir, cuando se estima el funcionamiento del ítem mediante una función matemática, Curva Característica del Ítem (CCI); o, bien, no paramétrica, que es cuando no se estima el funcionamiento del ítem a partir de una función matemática, sino a partir del funcionamiento observado del ítem. Las estimaciones del funcionamiento del ítem, en ambos casos, se aplican para los diferentes grupos en cada nivel del rasgo medido (Recio 2012).

La mayoría de las técnicas desarrolladas para detectar la presencia de DIF en un test han sido planteadas originalmente para su aplicación en ítems de respuesta dicotómica, es decir, aquellos ítems en que el sujeto evaluado debe elegir entre dos categorías de respuesta o, bien, en que sus posibilidades de respuesta pueden reducirse a

dos categorías. Sin embargo, gran parte de los test aplicados en la medición psicológica y, por extensión, en educación, posee ítems con un formato de respuesta que supera las dos categorías, es decir, posee patrones politómicos de respuesta, pudiendo estos ser nominales, cuando las posibles respuestas no están jerarquizadas, u ordinales, cuando cada posibilidad de respuesta adquiere un valor distinto, permitiendo su jerarquización.

Lo anterior ha derivado en el desarrollo de técnicas específicas para detectar el DIF en ítems de respuesta politómica. Recio (2012) propone una clasificación de estas técnicas, señalando que varias de ellas son generalizaciones de las técnicas aplicadas en casos de ítems dicotómicos, mientras otras se han desarrollado para ser aplicadas en cualquier tipo de ítems. (Figura n°3)

		Forma de caracterizar el funcionamiento del ítem	
		Paramétrico	No paramétrico
Forma de estimar el rasgo medido	Puntuación observada	Análisis Discriminante Logístico (Miller & Spray, 1993) Regresión Logística Politómica (Agresti, 1990; Frencg & Miller, 1996).	MH generalizado (nominal) (Mantel & Haenszel, 1959) Mantel (Mantel, 1963) Mantel-Haenszel Generalizado (Landis et al. , 1978)* Estandarización (Dorans & Schmitt, 1991) Pruebas Z (Welch & Hoover, 1993)
	Variable latente	Basados en la TRI	Medidas de área (Cohen, Kim & Baker, 1993) Test de Razón de verosimilitud (Kim & Cohen, 1998) Procedimiento DFIT (Flowers, Oshima & Raju, 1999)
		No basados en la TRI	Análisis Factorial Confirmatorio (Oort, 1992)
			Polytomous SIBTEST (Chang, Mazzeo & Roussos, 1996).

Figura n°3. Taxonomía de procedimientos para detectar DIF politómico. Adaptada de Recio (2012).*Agregado a la propuesta original.

Entre las técnicas que han surgido como adaptaciones de métodos para ítems dicotómicos, se pueden mencionar los métodos basados en la Teoría de Respuesta al Ítem (TRI) (Dorans, Schmitt, & Bleistein, 1992; Oshima & Raju, 1997; Raju, van der Linden, & Fler, 1995). También, existen adaptaciones de las técnicas basadas en análisis de regresión logística (Miller & Spray, 1993). Otras técnicas corresponden a la generalización del método de Mantel-Haenszel (Landis et al., 1978; Zwick & Thayer, 1996) y al método de Diferencia de Medias Estandarizadas (SMD), (Dorans & Schmitt, 1991; Dorans et al., 1992; Zwick & Thayer, 1996; Zwick et al., 1997). Por último, cabe mencionar dos procedimientos cuya aplicación sobre ítems politómicos no ha requerido de una adaptación especial. Estos son la técnica SIBTEST (Chang & Mazzeo, 1996) y el método basado en el Análisis Factorial Confirmatorio (Oort, 1992).

2.4 Revisión de los resultados de la Evaluación Docente y definición de variables para el análisis.

Al analizar los resultados de la Evaluación Docente, desde su implementación hasta el año 2010, se observan diferencias en el desempeño de los docentes que, en buena parte, pueden ser atribuidas a condiciones individuales, laborales y a las características del contexto (Alvarado & Cabezas, 2012; Bravo et al., 2008; Santelices et al., 2015; Santelices & Valenzuela, 2015; Taut et al., 2012; Valencia & Manzi, 2011). De este modo, año tras año, se consolida la tendencia de los resultados que favorecen a mujeres sobre hombres o, bien, a un determinado rango etario por sobre otros, entre otras diferencias. Sin embargo, frente a estas diferencias, se hace necesario evaluar la validez psicométrica de las mismas para descartar que sean el resultado de una distorsión aportada por el propio instrumento de medición, sobre todo cuando dichas diferencias parecen multiplicarse y/o acentuarse cuando se observan los mismos resultados desagregados al nivel de cada indicador medido en la Evaluación Docente.

2.4.1 Sexo

De acuerdo con los estudios realizados sobre resultados entre los años 2005-2010, las profesoras constituyen la mayoría de la dotación docente con que cuenta el sistema escolar municipal. Al mismo tiempo, una proporción del 64% de ellas logra un desempeño Competente o Destacado, contra un 57% de los hombres, lo que se traduce en una mayor probabilidad de obtener un buen desempeño en la Evaluación Docente para las mujeres. (Valencia & Manzi, 2011).

Los resultados correspondientes al año 2013 son consistentes con la tendencia mencionada, mostrando que se mantiene la mayor proporción de mujeres (71,7%) que obtienen un desempeño Competente o Destacado con respecto a los hombres con el mismo desempeño (63,2%) (MINEDUC, 2013). Al mismo tiempo, los resultados dan cuenta de que las mujeres no solo alcanzan un mejor resultado, sino que la tendencia en el aumento de la proporción para los niveles agrupados Competente y Destacado crece en magnitudes similares para ambos sexos. Lo mismo sucede con la disminución de la proporción para el nivel Insatisfactorio: la proporción de mujeres en esta categoría se reduce en -1,4% (de 2,9% a 1,5%) mientras la proporción de hombres varía en -2,6%. (de 5,5% a 2,9%). Las mujeres no solo muestran una mayor probabilidad de obtener un buen desempeño, sino también presentan un 7,1% mayor en la probabilidad de mejorar sus puntajes, 35,6% contra 28,5%; y probabilidad más baja de disminuir sus puntajes: 17,5% versus el 25,4% que presentan los hombres (MINEDUC, 2013).

A lo anterior se agrega que, en nivel de Primer Ciclo, el porcentaje que mejora sus puntajes tanto entre mujeres como en hombres es mayor que el porcentaje que los disminuyen, sin embargo, las profesoras vuelven a mostrar cierta ventaja por sobre los varones; en el caso de las mujeres, mejora su puntaje un 45% versus un 37,7% que los disminuye; y, en el caso de los hombres, mejora su puntaje un 43,6% versus un 29,2% que los disminuye. En Segundo Ciclo de enseñanza básica, las mujeres también mejoran, pero los hombres retroceden en cuanto a la proporción de los que empeoran. En

Enseñanza Media, por su parte y aunque levemente, es mayor la proporción de docentes que empeora sus puntajes que la que los mejora y su comportamiento es muy similar para hombres y mujeres (MINEDUC, 2013).

Al comparar la evolución de los resultados en el instrumento portafolio, se observan diferencias a favor de las mujeres en los grupos iniciales extremos: Insatisfactorio y Destacado. (MINEDUC, 2013)

En el segmento Insatisfactorio inicial, en un entorno de mejora en todas las dimensiones del Portafolio, las mujeres muestran mayor progreso relativo que los hombres en las dimensiones Organización de la unidad y Análisis de la clase: aumento de 0,65 versus 0,60 puntos y Análisis de la clase, con 0,64 puntos de aumento versus 0,61.

En el segmento Destacado inicial, en un entorno de disminución en seis de las siete dimensiones del Portafolio y de menos 0,38 puntos en su puntaje promedio, los hombres muestran una menor caída relativa que las mujeres en la dimensión Ambiente de la clase: con una disminución de -0,11 versus -0,16 de los mujeres (MINEDUC, 2013).

Resulta pertinente señalar que, para la variable “sexo”, se llevó a cabo un análisis preliminar sobre los datos del Portafolio 2009. Dicho análisis utilizó el procedimiento de Mantel–Haenszel para ítems politómicos y análisis de invarianza de medición, sin que se reportara presencia de DIF en contra del grupo focal (hombres) (Taut et al., 2011). Sin embargo, resulta pertinente mencionar que la publicación no especifica si los análisis se realizaron en función de las dimensiones o, bien, a partir de cada uno de los indicadores del portafolio.

2.4.2 Edad

Las investigaciones que buscan establecer relaciones entre la edad de los docentes y la calidad de su desempeño profesional no son totalmente concluyentes. Una revisión de la literatura internacional realizada por Hunt (2009) señala al respecto que la relación entre esta variable es irrelevante. Por otro lado, a nivel nacional, cabe

mencionar el estudio realizado por Eisenberg (2008), con datos de docentes generalistas de la Evaluación Docente 2005, donde utiliza dos niveles que consideran, en primer término, conjuntamente características individuales y de conjunto (docente – escuela) y, en un segundo nivel, solo aspectos del contexto (comuna). Esta investigación analizó la relación entre la experiencia profesional, altamente correlacionada con la edad, y el desempeño docente. A partir de los resultados del mencionado estudio, se concluyó la existencia de una relación inversa entre las variables edad y desempeño en la evaluación docente (Eisenberg, 2008). Esta evidencia contrasta con lo aportado por Bravo y Falck (2008), quienes, en un análisis realizado sobre los datos de la Evaluación Docente, concluyen que la relación entre la edad y el desempeño docente es positiva, pero decreciente en el tiempo (Bravo et al., 2008).

Otras investigaciones señalan que la relación entre la edad y el desempeño docente es una función cuadrática cuya curva describe la forma de U invertida, en dicha curva, los docentes más jóvenes presentan una pendiente ascendente, alcanzando el desempeño óptimo entre los 41 y 45 años para, luego, presentar una caída progresiva (Sturman, 2003; Valencia & Manzi, 2011). No se encontraron estudios que analicen la presencia de DIF a partir de esta variable en ninguno de los instrumentos de la Evaluación Docente.

2.4.3 Zona Geográfica (rural/urbana)

Entre los años 2005 y 2010, un 29% de los docentes evaluados por el Sistema Nacional de Evaluación Docente se desempeñaba en zonas rurales, mientras el 71% restante lo hacía en escuelas emplazadas en zonas urbanas. Los resultados comparados según la zona geográfica donde se ubica la escuela (rural/urbana) arroja un 5% de diferencia en la proporción de docentes Competentes y Destacados a favor de los docentes que trabajan en contextos urbanos (63%), sobre aquellos que lo hacen en zonas rurales (58%) (Valencia & Manzi, 2011).

Es importante señalar que las particularidades de las escuelas rurales constituyen un marco de interacciones distintas entre docentes y estudiantes respecto de las realidades presentes en los contextos urbanos. A diferencia de las zonas urbanas, en el contexto rural, existen escuelas uni, bi, tridocentes y polidocentes. Muchas de estas escuelas se concentran en zonas pobres y con pocas posibilidades de desarrollo, así mismo, otras funcionan en zonas económicamente productivas, donde la modernización de la industria agrícola y ganadera demanda a las escuelas niveles escolarización capaces de responder a las exigencias y necesidades propias de estas actividades económicas (Gajardo & De Andreaca, 1992; Vera et al., 2012; Williamson, 2004). Otro dato importante es que la mayor parte de las escuelas rurales pertenece a la administración municipal (81%), en tanto que un 18,4% corresponde al sector particular subvencionado (San Miguel, 2005; Vera et al., 2012; Williamson, 2004).

Según un estudio de Feu i Gelis (2004), las características y necesidades del contexto rural permiten esperar que en estas escuelas se favorezca la experimentación educativa y el desarrollo de una pedagogía activa, puesto que los docentes que se desempeñan en escuelas rurales pueden impartir una enseñanza individualizada en un contexto diverso, con un sistema de enseñanza más flexible que en contextos urbanos (Feu i Gelis, 2004; Vera et al., 2012). Las necesidades de adaptarse a las particularidades del medio rural, para el cual muy pocos docentes han recibido preparación especial, exige a los profesores de estas escuelas la adecuación de sus prácticas pedagógicas según las condiciones y necesidades de este medio educativo (Vera et al., 2012; Villarroel et al., 2003).

Otras investigaciones se centran en las expectativas que, de acuerdo con las necesidades y condiciones de los sectores rurales, deben plantearse respecto al desempeño docente en dichos contextos ya que, dadas las condiciones actuales, nos encontramos en las escuelas rurales con un cuerpo docente heterogéneo en cuanto a sus características personales y trayectoria profesional, enfrentados a condiciones también

variables de acuerdo a las características de la zona en que se ubica la escuela (Vera et al., 2012). Lo anterior evidencia la necesidad de definiciones menos difusas en cuanto a lo que se espera del docente y cómo su desempeño se entiende integrado a una auténtica participación comunitaria que excede ampliamente la relación profesor-alumno (Thomas & Hernández, 2005; Vera et al., 2012; Villarroel et al., 2003). En esta misma perspectiva, algunas visiones señalan que la variedad de funciones y roles que demanda el quehacer docente en contextos rurales puede afectar el desempeño de los profesores, incidiendo negativamente en factores críticos para la efectividad de la enseñanza como la motivación para enseñar (Schiefelbein, 1992; Thomas & Hernández, 2005).

Por otra parte, las relaciones pedagógicas que tienen lugar en contextos rurales se sustentan en percepciones de un entorno, cuyas características específicas afectan el comportamiento docente, dichas percepciones, siguiendo la propuesta de Vera (2012), son filtradas por las creencias del docente, determinando la forma en que este conduce el proceso educativo dentro del aula (Prieto & Contreras, 2008; Vera et al., 2012). Las creencias pueden incorporarse al proceso de enseñanza a través de las expectativas que el profesor tiene sobre el desempeño del estudiante (Fang, 1996), lo que resulta relevante al revisar la evidencia que asocia positivamente altas expectativas sobre los estudiantes, con mejores prácticas docentes y mayor efectividad en la enseñanza (OCDE, 2009; Santelices & Valenzuela, 2015; Valencia & Manzi, 2011).

La relación entre expectativas de los docentes sobre el desempeño de sus estudiantes y el desempeño en la Evaluación Docente ha sido analizada por Valencia y Manzi (2011), revelando que aquellos docentes que creen que sus estudiantes terminarán la educación media o lograrán estudios superiores obtienen un mejor desempeño que los docentes que declaran expectativas más bajas de sus estudiantes. La proporción de docentes que obtiene un desempeño Competente o Destacado entre quienes declaran altas expectativas de logro de sus estudiantes, es decir quienes creen que sus estudiantes terminarán la Enseñanza Media, es de un 63%. Dicha proporción baja a un 53% cuando la expectativa es Enseñanza Media incompleta y 46% para Enseñanza Básica completa

(Valencia & Manzi, 2011).

Nicole Eisenberg (2008) analizó los datos de 5.164 docentes generalistas evaluados el año 2005 en la Evaluación Docente. En dicho estudio se aplicó una metodología multinivel, considerando 2 niveles, para explicar los resultados de desempeño de los profesores. En el primer nivel, se incluyen en forma conjunta antecedentes del nivel individual de los docentes y la escuela; y, en un segundo nivel, antecedentes de la comuna (contexto) (Valencia & Manzi, 2011). Es interesante señalar que los hallazgos de esta investigación señalan que la disminución de los resultados en los docentes de mayor edad se acentúa en contextos de ruralidad. Sin embargo, no se encontró una asociación directa entre el desempeño docente y el tipo de zona geográfica donde se ubica la escuela (rural/urbana) (Eisenberg, 2008; Valencia & Manzi, 2011).

En cuanto a la posibilidad de que las diferencias en los resultados de la Evaluación Docente relativas a la variable de la zona geográfica (rural-urbana) se deban a características psicométricas de los instrumentos de medición aplicados, el informe Evaluación Docente en Chile (Manzi, González, & Sun, 2011) reporta un estudio realizado sobre los resultados del portafolio 2009 analiza la presencia de DIF utilizando la metodología Mantel-Haenszel para comparar el funcionamiento de los ítems entre docentes de escuelas rurales y urbanas, sin encontrarse evidencia de sesgo en contra del grupo focal (escuelas rurales). Sin embargo, el mismo estudio señala la necesidad de profundizar en estos análisis (Valencia & Manzi, 2011).

2.4.4 Nivel de Enseñanza y Asignatura

Cuando se analizan los resultados de la Evaluación Docente según la asignatura y el nivel de enseñanza, se observan marcadas diferencias en la proporción de docentes que alcanzan los niveles de desempeño Competente y Destacado. A modo de ejemplo, al observar los resultados de la Evaluación Docente entre los años 2006 y 2009, en el

indicador Secuencia de Clases, existe un rango de variación entre los grupos evaluados desde el 11,8% al 73,1%, presentándose la más alta proporción de docentes con nivel Competente o Destacado en Artes Musicales para Educación Media y la más baja proporción en la asignatura de Física, donde sólo un 11,8% obtiene un buen resultado. Esta diferencia se replica en otros indicadores medidos en la Evaluación Docente (Sun, Correa, et al., 2011). Frente a esta evidencia, es pertinente señalar que los análisis inferenciales aplicados en el presente trabajo, sobre los resultados del portafolio 2013, muestran que el desempeño general de los docentes de las asignaturas de Artes Musicales y Física en el instrumento portafolio se acerca al promedio del puntaje global. Lo anterior resulta especialmente interesante cuando observamos que la estructura del portafolio fue modificada el año 2012, por lo que el indicador al que se hace referencia en Sun, Correa, et al. (2011) no está presente en la versión del portafolio utilizada para el presente estudio.

Otras investigaciones también han indagado en la influencia del nivel de enseñanza y la asignatura, y los resultados de la Evaluación Docente, en esta perspectiva, Valencia (2009, citado en Valencia y Manzi, 2011) señala que los docentes generalistas estuvieron asociados a los resultados más bajos, mientras que aquellos de Lenguaje y Matemáticas de Segundo Ciclo y los de Historia en Enseñanza Media se asociaron a un desempeño más alto.

Otras evidencias son aportadas por un estudio realizado con datos de 1.422 docentes de 4° u 8° básico, evaluados entre 2006 y 2008. Esta investigación utilizó una estructura de tres niveles que consideró diferencias individuales, entre escuelas y entre comunas. Un hallazgo importante señala que una proporción del 24% del desempeño docente se explica por antecedentes del contexto (9%, atribuida a diferencias en la escuela; 15%, a la comuna). Sin embargo, se destaca que la parte atribuible al contexto varía según el nivel de enseñanza considerado (Taut, Santelices, & Valencia, 2010; Valencia & Manzi, 2011).

En suma, luego de revisar la evolución de la Evaluación Docente desde su implementación hasta la fecha, proceso que ha alcanzado cobertura universal en el sistema escolar municipal, parece evidente la importancia de este proceso, cuyas implicancias alcanzan hoy una trascendencia que la pone al centro de profundos debates en el contexto de definiciones de un programa de reformas educativas. El nuevo escenario supone una serie de desafíos para todos los actores involucrados en el sistema escolar, donde, por cierto, los profesores juegan un papel fundamental. Por esta razón, las políticas públicas orientadas a fortalecer la profesión docente, en sus distintas dimensiones, constituyen un pilar esencial, donde un sistema de evaluación docente aparece como una herramienta central para dotar de información acerca de las verdaderas fortalezas y debilidades de los profesores, permitiendo canalizar los esfuerzos de forma eficiente, lo que asegura que el sistema escolar cuente con profesionales a la altura de las necesidades de educación chilena.

Dadas las características y consecuencias asociadas a estos procesos de evaluación de desempeño, resulta de suma importancia que los instrumentos de medición funcionen con los mayores niveles posibles de estabilidad y validez, evitando que la interpretación de los resultados pudiese verse afectada por distorsiones surgidas a partir de las características psicométricas de los mismos instrumentos. Esto, pues lo esperable es que, en una medición sobre una muestra en la que los sujetos presentan diferentes características socio-demográficas o grupos de pertenencia, tales como edad, sexo, región de origen, etc., los sujetos que presenten el mismo nivel del atributo medido presenten resultados idénticos, sin importar su grupo de pertenencia. En el caso de que uno o más de los ítems que componen el instrumento de medición entregue resultados diferentes (a favor o en contra) de sujetos que tienen el mismo nivel de habilidad o atributo medido, se habla de funcionamiento diferencial del ítem (DIF, Differential Item Functioning), término introducido por Holland y Thayer (Holland & Thayer, 1986, 1988).

Con el objeto de ofrecer un panorama más claro acerca de las problemáticas abordadas en el presente capítulo, se presenta una breve síntesis de los temas desarrollados, y sobre los cuales se articula el presente trabajo.

En primer lugar, se hace referencia a la discusión, tanto a nivel nacional como internacional, en torno a las conceptualizaciones de la calidad docente y los diferentes acercamientos metodológicos para determinar mediciones de la misma, que permitan orientar la toma de decisiones en materia educativa.

Un segundo tema desarrollado en este apartado se relaciona con el sistema de Evaluación Docente adoptado por el sistema escolar público de Chile. Se describen los diferentes instrumentos evaluativos utilizados, profundizado en las características del portafolio, y en las implicancias de sus resultados, los cuales se utilizan para llevar a cabo los análisis implementados en el presente trabajo.

Un tercer tema fundamental de este estudio, se refiere a la definición del Funcionamiento Diferencial de Ítem (DIF) y la importancia de su detección para garantizar la validez de los procesos evaluativos, especialmente cuando se consideran las altas consecuencias que tiene la Evaluación Docente para los el sistema educativo chileno. Del mismo modo, se aborda una revisión de diferentes procedimientos que se han desarrollado para el estudio del DIF.

Un cuarto tema que articula este capítulo, corresponde a las diferencias que se han reportado en los resultados del portafolio, diferencias que pudieran estar asociadas a diferentes variables, tales como *Sexo, Edad, Zona Geográfica, Nivel de Enseñanza y Asignatura* impartida por el docente evaluado. A partir de la existencia sistemática de dichas diferencias, se fundamenta la implementación de los análisis DIF, con el motivo de determinar si dichas diferencias se explican o no por las características psicométricas del instrumento portafolio.

III Marco Metodológico

En el presente capítulo se revisará cada una de las etapas contempladas en la estrategia metodológica diseñada en función de los objetivos declarados para el presente estudio. Esta estrategia se sustenta dentro del paradigma de investigación cuantitativa, basada en el análisis secundario de datos por medio de procedimientos estadísticos, con el fin de determinar, en una primera etapa, la existencia de diferencias significativas en los resultados reportados por cada indicador del portafolio de la evaluación docente, y en una segunda etapa, determinar si dichas diferencias en los resultados son atribuibles a las características psicométricas del instrumento portafolio.

3.1 Población y muestras

La población considerada en el presente estudio fue la cohorte 2013 de la Evaluación Docente. Esta cohorte corresponde a 17.070 docentes empleados en el sistema de educación municipal chileno a lo largo de todo el país, correspondientes al 19,65% del universo total de docentes pertenecientes al sistema escolar municipal, cuyo número al año 2013 registraba 86.858 docentes. De los docentes que rindieron la evaluación ese año, un 72,8% corresponde a mujeres, mientras un 27,2% son de sexo masculino. La Edad promedio de los evaluados es de 42 años; por otro lado, un 75,6% de la población en estudio se desempeña en establecimientos urbanos, mientras un 24,4% lo hace en contextos rurales. En cuanto al nivel de enseñanza que los docentes atienden, éstos se distribuyen en 6 grupos: Educación Parvularia [EP] (6,8%), Primer Ciclo Básico [PC] (25,9%), Segundo Ciclo Básico [SC] (32,8%), Educación Media [EM] (20,4%), Educación Especial [EE] (12,3%) y Educación de Adultos [EA] (1,8%), concentrando la mayor proporción de docentes el grupo que corresponde a SC, y la menor proporción el grupo EA. La asignatura por la que cada docente es evaluado se encuentra estrechamente vinculada al nivel de enseñanza atendido por el profesor, de modo que un docente de PC no puede ser evaluado por una asignatura presente en los

planes de estudio de EM, y viceversa, por esta razón, la distribución de docentes correspondiente en función de las 31 asignaturas contempladas por la evaluación docente, será tratada con mayor detalle en el apartado *Variables*, del presente capítulo, a partir de la relación Nivel de Enseñanza-Asignatura.

3.2.1 Base de datos y su preparación

La base de datos para realizar el presente estudio fue proporcionada por el Centro de Estudios del Ministerio de Educación, previa suscripción de un compromiso para la utilización de la información con fines exclusivamente académicos, y el resguardo por la confidencialidad de los datos.

Como primer paso para preparación de la base de datos, se descartaron los casos que presentaban valores perdidos para alguna de las variables a analizar. Este procedimiento se realizó previo análisis de valores perdidos, el cual arrojó que dichos valores representaban una proporción muy baja para todas las variables en estudio, siendo Edad la variable que presentó un mayor porcentaje de valores perdidos (5,4%). Sin embargo, las estimaciones arrojadas por el análisis de valores perdidos, no se aprecian diferencias estadísticamente significativas en la media, como en la desviación estándar de la variable edad en cuatro opciones de tratamiento de estos valores: Por lista, Todos los valores, EM y Regresión. Adicionalmente a los valores perdidos en esta variable, se eliminaron 34 casos con valores considerados inverosímiles, puesto que señalaban edades menores a 22 años.

A los valores descritos, se agregan 8 casos sin puntaje asociado a los indicadores del módulo N°2, codificados con ausencia de la grabación, o falta de parte ella, por problemas atribuibles al camarógrafo que recogió la evidencia, estos casos también fueron excluidos de la base de datos por representar una proporción marginal de la misma.

Una vez depurada, la base de datos obtenida contó con 16.114 casos, correspondiente al 94,4% de la población de docentes evaluados, de los cuales un 73% corresponde a mujeres y un 27% a hombres; el 75,5% corresponde a docentes que se desempeñan en escuelas urbanas, contra un 24,5% que lo hace en contextos rurales. Las proporciones en que se distribuyen los docentes según nivel de enseñanza y asignatura tampoco presentan variaciones significativas.

3.2.2 Extracción de Muestras

TAMAÑO DE LA MUESTRA

No existe un consenso definitivo frente a la problemática del tamaño muestral necesario para implementar análisis DIF. Algunos estudios sugieren un tamaño mínimo de 1000 casos para implementar análisis estadísticos orientados a determinar la presencia de DIF (S.-H. Kim & Cohen, 1997; S.-H. Kim, Cohen, & Alagoz, 2007), sin embargo existen algunas consideraciones que deben tenerse en cuenta.

En los procedimientos MH, la tasa de Error tipo I (falsos positivos) aumenta cuando aumenta el tamaño de la muestra (Wu & Tsai, 2010). Lo anterior se explica porque la potencia estadística de la prueba es una función del tamaño de la muestra (Cohen, 1988; S.-H. Kim et al., 2007; Narayanan & Swaminathan, 1994), y al aumentar la potencia estadística, el procedimiento se vuelve más susceptible a identificar asociaciones significativas.

Por otra parte, otros estudios señalan que muestras de mayor tamaño entregan resultados más estables (Zilberberg, Phan, Socha, Kong, & Keng, 2002; Zwick, 2012). Cuando el tamaño de la muestra del grupo focal es mayor, la tasa de DIF detectado erróneamente aumenta (Aguerri, Galibert, Zanelli, & Attorresi, 2005). El ETS señala que el mínimo tamaño de la muestra para análisis DIF, aplicados test en etapa de ensamblaje, es de 200 miembros en el grupo más pequeño, y 500 en la muestra total; y

un mínimo de 300 miembros en el grupo más pequeño y 700 en total, cuando el análisis se aplica sobre test que ya fueron aplicados, pero previo a que los resultados hayan sido reportados (Zwick, 2012). Dado que los análisis DIF forman parte de los exámenes para evaluar la validez de los ítems de un test.

A pesar de que los ítems de respuesta abierta (constructed response), forman parte de los estudios del ETS desde hace casi dos décadas, éste no ha logrado adoptar un criterio definitivo en relación al tamaño de la muestra para el análisis de este tipo ítems, siendo este un tema aún no resuelto en cuánto a los análisis DIF sobre este tipo de ítems. Algunas de las dificultades que han enfrentado dicen relación con la aplicación rutinaria de análisis DIF para este tipo de test es baja (Moses, Liu, Tan, Deng, & Dorans, 2013).

No obstante el presente estudio se implementa sobre los resultados de un proceso de evaluación ya aplicado, se ha adoptado el criterio sugerido por el ETS para instrumentos evaluativos en etapa de ensamblaje; es decir, un mínimo de 200 miembros para el grupo más pequeño, y un mínimo de 500 casos para ambos grupos sumados. Lo anterior se fundamenta en el hecho de que un criterio más exigente obligaría a excluir del análisis algunas comparaciones pareadas de interés, entre grupos que han presentado diferencias significativas en los resultados históricos reportados para este instrumento. Se asume, por tanto, que no habiendo un criterio único y definitivo en materia del tamaño muestral mínimo para la aplicación de análisis DIF, resulta conveniente adoptar un criterio menos riguroso con el fin de obtener datos más informativos como resultado del presente estudio.

A partir de la base depurada, se extrajeron 9 muestras diferentes para llevar a cabo los análisis. Para evaluar el DIF, aplicando los procedimientos generalizados MH, en las variables Sexo, Zona Geográfica, Edad (por rangos) y Nivel de enseñanza, se utilizaron muestras de 1000 casos, extraídas al azar para cada análisis. Además, se implementó para estas variables un segundo análisis utilizando una única muestra, con

10000 casos extraídos al azar.

La variable asignatura se evaluó comparando docentes dentro de un mismo nivel de enseñanza, razón por la cual se extrajeron muestras que incluyeron a todos los docentes pertenecientes al respectivo nivel de enseñanza. De este modo se obtuvo la Muestra SC corresponde a los 5.303 casos correspondientes al nivel de enseñanza Segundo ciclo. Al igual que en el caso anterior, la Muestra EM concentra los 3.286 casos correspondientes al nivel de enseñanza Educación Media, la muestra EE a los 1.986 casos del nivel Educación Especial, y la Muestra EA a los 278 casos correspondientes al nivel de enseñanza Educación de Adultos. No se extrajeron muestras para los niveles Educación Parvularia y Primer Ciclo dado que en dichos niveles no hay diferencias por asignatura entre los docentes evaluados dentro de la misma cohorte.

3.3 Consideraciones metodológicas acerca del instrumento de medición

Una primera consideración metodológica a tener presente, es que, para efectos de este trabajo y todos los procedimientos de análisis aquí implementados, el portafolio será considerado un test con puntuaciones iguales para cada indicador, entendiendo el estudio DIF como un análisis de las propiedades psicométricas de cada indicador, independiente del proceso de ponderación que deriva en la determinación del puntaje final del instrumento portafolio en el contexto del proceso de Evaluación Docente. De este modo, los análisis consideran variables de emparejamiento definidas por la sumatoria de puntajes obtenidos en conjuntos de indicadores, sin considerar la ponderación que implica la pertenencia a una dimensión u otra, como tampoco se vincula con las decisiones para determinar los puntajes de corte que determinan finalmente los niveles de desempeño reportados para este instrumento. De todos modos, con el fin de determinar si la utilización de los puntajes sumados modificaría significativamente la distribución de la población, pudiendo conducir a una interpretación distorsionada del desempeño docente, se analizó la relación entre la variable que contenía la puntuación

final en el instrumento (ponderando el peso de cada indicador según su dimensión de pertenencia) “Prom_Pond_Portafolio” y una nueva variable calculada con la sumatoria de los puntajes en cada indicador, “Suma_Portafolio”, encontrándose una correlación fuerte entre ambas variables ($r=0,986$).

Un segundo tema a tener en consideración, es la estructura dimensional del instrumento portafolio, puesto que los procedimientos de análisis acá implementados, utilizan variables de emparejamiento definidas por la puntuación obtenida en el test completo, lo que significa que todos los ítems del test están enfocados en la medición de una única dimensión. Cuando se viola el supuesto de la unidimensionalidad, las estimaciones basadas en la comparación entre las proporciones de respuesta de cada ítem y las proporciones de respuesta al test completo, pueden conducir a sacar conclusiones erróneas acerca de los ítems. Lo anterior significa que si un evaluado tiene diferentes niveles de habilidad en los diferentes rasgos medidos, estos rasgos pueden aparecer sobre representados o sub representados en el puntaje total del test, lo que podría conducir a estimaciones equivocadas acerca del comportamiento de los ítems que miden dichos rasgos. Por esta razón, los procesos cuyas variables de emparejamiento se obtienen a partir del puntaje total, requieren que se cumpla el supuesto de la unidimensionalidad, lo que significa que las respuestas serán producidas en base a un único atributo o rasgo latente (Hattie, 1985). A partir de lo anterior, cabría esperar que la varianza de un test fuera explicado por una única variable latente, lo que es ciertamente improbable en la realidad (Wright & Linacre, 1989), por lo que se han definido una serie de criterios para determinar si un test cumple con el mínimo de unidimensionalidad necesaria, que permita considerar que mide un único constructo. Para el presente estudio, se considerará el criterio de Reckase (1979), según el cual, un test se considera unidimensional si el primer factor extraído explica un 20% de la varianza total del test.

No obstante lo anterior, un test puede estar diseñado multidimensionalmente, es decir, se espera que el test mida variables que representan subescalas o dimensiones de

un constructo más amplio, o bien, que para su respuesta requieran de procesos cognitivos diversos. En la práctica, es esperable que, en contextos de evaluación educativa, un test mida más de una dimensión, es decir, que en algún grado, no se cumpla el supuesto de la unidimensionalidad (Reckase, 1979). La violación sistemática de este supuesto, por razones cuya legitimidad no está en entredicho, genera una problemática compleja en relación con los análisis DIF, Fidalgo y Ferreres (2002) señalan que, cuando un test busca medir diferentes dimensiones, no debe utilizarse el puntaje final del test para estimar el funcionamiento de un ítem, sino que debe definirse una variable de emparejamiento a partir de la suma de las puntuaciones de los ítems de cada dimensión. Por otra parte, una postura algo diferente, sostiene que en los análisis DIF, el desempeño de cada ítem se evalúa en relación con un constructo particular, cuya medición está controlada por la puntuación del test (Linn & Werts, 1971), en virtud de esto, a pesar de reconocer en la dimensionalidad un aspecto en el cual se debe profundizar las investigaciones, y que este aspecto puede constituir una posible fuente de complicaciones para los análisis DIF, estas complicaciones parecen despreciables frente a otras influencias que pueden socavar los estudios DIF (Dorans, 2013).

En cuanto a la unidimensionalidad del portafolio, es esperable que, dados los fines que persigue, y considerando su diseño, no cumpla con dicho supuesto, esto se acentúa si consideramos, además, la diferencia en el tipo de información recogida por los dos módulos que componen su estructura más gruesa, mientras el Módulo1 recoge información y evidencia documental del quehacer docente, el Módulo2 recoge información por medio de un registro performático, la realización de una clase, lo que implica que los procesos cognitivos involucrados para el desarrollo de ambos módulos involucran habilidades y competencias diferentes. Por otra parte, la posibilidad de considerar subescalas o sustest a partir de cada módulo, o de cada una de las dimensiones reviste una dificultad técnica para la implementación de los análisis DIF, dado que el reducido número de ítems presentes en cada dimensión del portafolio dificulta los procedimientos de análisis estadísticos. Además, al tratar cada dimensión

como un test independiente, es esperable que la consistencia interna de cada dimensión disminuya, dada la sensibilidad de los coeficientes alpha a la disminución de la cantidad de ítems.

3.4 Operacionalización de las variables en estudio

3.4.1 Sexo

Esta variable corresponde a la variable Género del Docente, se procesó manteniendo su codificación en la base de datos original, correspondiendo el valor 1 para el sexo masculino, y 2 para el sexo femenino. Para objeto de los análisis, el grupo correspondiente al valor 2 (femenino), será considerado como grupo de referencia, mientras 1 (masculino) corresponde al grupo focal.

3.4.2 Zona geográfica (Rural/Urbano):

Esta variable también utiliza una codificación de dicotómica, asignándose el valor 0 al grupo de referencia, formado por las 12.169 escuelas emplazadas en zonas urbanas (75,5%), y el valor 1 a las 3.945 escuelas de zonas rurales (24,5%), correspondientes al grupo focal.

3.4.3 Edad (por rangos):

Esta variable corresponde a la recodificación de la variable Edad, para la cual se definieron rangos de 5 años, en concordancia con los reportes de resultados y los estudios que han indagado posibles interacciones de esta variable. De este modo, se agrupó esta variable en 8 rangos etarios, quedando definida como una variable ordinal con los siguientes valores: 1= menos de 30 años (22,7%); 2=31 a 35 años (14,9%); 3=36 a 40 años (10,8%); 4=41 a 45 años (9,3%); 5=46 a 50 años (12,3%); 6=51 a 55 años (14,1%); 7=56 a 60 años (13,0%); 8=61 años o más (3,1%). A pesar de que la literatura señala que la mayor proporción de docentes con altos niveles de desempeño se ha concentrado históricamente en el rango etario entre los 41 y 45 años (Manzi, 2011), para el presente estudio se considerará como grupo de referencia al N°3, dado que los análisis muestran que el año 2013 los mayores resultados en el instrumento portafolio se concentraron entre los 36 y 40 años (*Figura n°4*). Sólo se realizarán comparaciones entre el grupo de referencia y los grupos que muestren diferencias estadísticamente significativas respecto de éste.

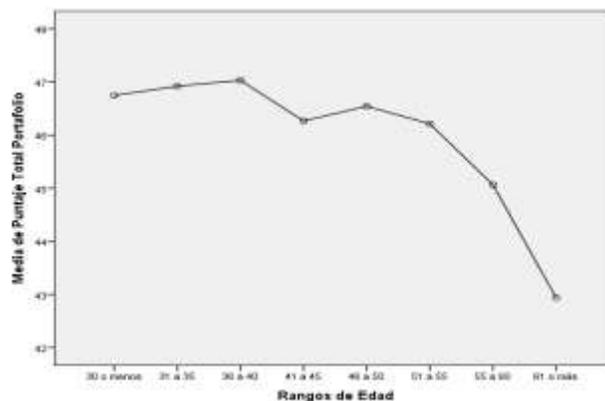


Figura n°4. Comparación de medias para cada rango de edad en la puntuación total del portafolio.

3.4.4 Nivel de Enseñanza:

Para el análisis de esta variable se codificaron 6 grupos correspondientes a los diferentes niveles de enseñanza en los que fueron evaluados los docentes, los valores numéricos utilizados para esta codificación tienen carácter nominal, y no corresponden a una jerarquización de estos niveles, puesto que responden únicamente a la necesidad de facilitar la operacionalización de la variable en el software GMHDIF.

Los valores asignados a los diferentes Niveles de Enseñanza, y su representación dentro de la muestra N°1 fueron los siguientes:

1=Educación Parvularia (EP), con 691 casos, correspondientes al 6,9% de la muestra; 2=Primer Ciclo Básico (PC), representando una proporción del 25,8% (2578 casos) de la muestra; 3=Segundo Ciclo Básico (SC), con un 33,5% (3354 casos) de la muestra; 4=Educación Media (EM), correspondiente a un 19,9% de la muestra (1986 casos); 5=Educación Especial (EE), que equivale al 12,3% de la muestra (1228 casos); y 6=Educación de Adultos (EA) con una proporción del 1,6% de la muestra, con 163 casos.

Al comparar el desempeño por Nivel de enseñanza, en función del puntaje total obtenido, se observa un mejor desempeño en los grupos PC y EE (*Figura n°5*).

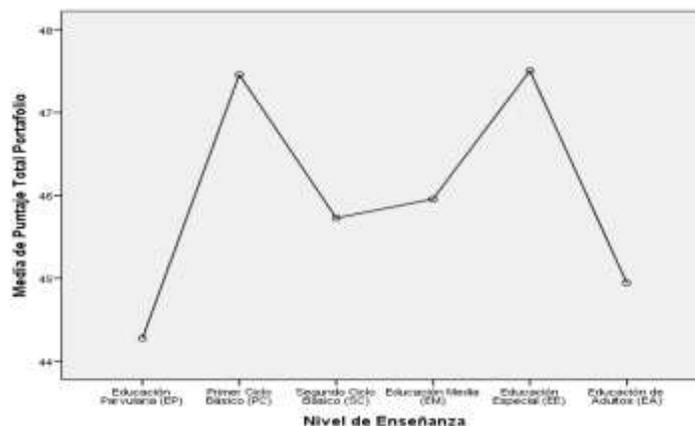


Figura n°5. Comparación de medias para la variable *Nivel de Enseñanza*

3.4.5 Asignatura:

En primer lugar, es necesario señalar que esta variable fue considerada en su relación con la variable Nivel de Enseñanza, por lo que cada asignatura fue definida como un grupo de comparación dentro de un determinado nivel de enseñanza. De este modo, se buscaba observar el comportamiento de los ítems, en relación con la variable *Asignatura*, entre docentes que atienden a cursos con estudiantes del mismo nivel, cuando dicho nivel presenta alternativas de asignaturas sobre las cuales desarrollar el instrumento de evaluación. Por esta razón, se excluyó del análisis para esta variable a los niveles Educación Parvularia y Primer Ciclo Básico. También se excluyó de este análisis el nivel Educación de Adultos por no satisfacer el criterio de tamaños muestrales mínimos para llevar a cabo los procedimientos estadísticos.

A partir de la base de datos previamente tratada se extrajeron aleatoriamente muestras específicas para cada uno de los niveles a analizar: Muestra SC, Muestra EM, Muestra EE y Muestra EA. En cada muestra se seleccionaron todos los casos correspondientes al nivel respectivo. Las diferentes asignaturas mantuvieron la codificación de la variable SUBSECT (subsector asociado al rut) presente en la base de datos original.

Las características de cada muestra, así como la representación porcentual de cada asignatura dentro la respectiva muestra fue la siguiente:

Muestra SC:

Cuenta con 5.303 casos distribuidos en 11 asignaturas, Artes Musicales [AM] (194 casos, 3,7%); Artes Visuales [AV] (176 casos, 3,3%); Ciencias Naturales [CN] (617 casos, 11,6%); Historia, Geografía y Ciencias Sociales [CS] (591 casos, 11,1%); Educación Física [EF] (617 casos, 11,6%); Educación Tecnológica [ET] (93 casos,

1,8%); Inglés [I] (503 casos, 9,5%); Lenguaje y Comunicación [L] (1056 casos, 19,9%); Matemática [M] (995 casos, 18,8%); Religión Católica [RC] (331 casos, 6,2%); y Religión Evangélica [RE] (130 casos, 2,5%). Ver *Figura 6*.

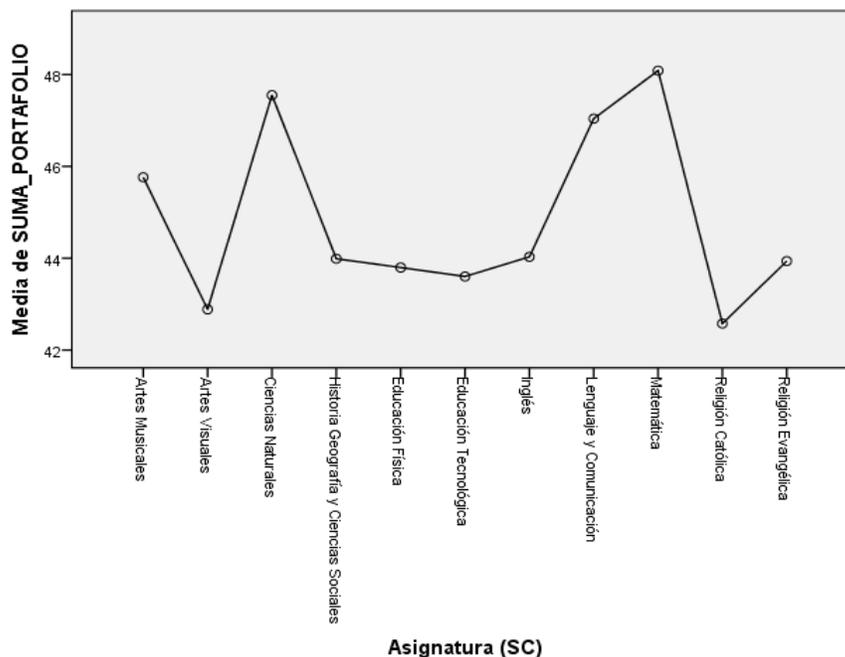


Figura n°6. Comparación de medias en la variable *Asignatura (Segundo Ciclo)*

Muestra EM:

Esta muestra cuenta con 3.286 casos distribuidos en 14 asignaturas, Artes Musicales [AM] (193 casos, 5,9%); Artes Visuales [AV] (239 casos, 7,3%); Biología [B] (179 casos, 5,4%); Educación Física [EF] (261 casos, 7,9%); Educación Tecnológica [ET] (167 casos, 5,1%); Física [F] (95 casos, 2,9%); Filosofía y Psicología [FP] (138 casos, 4,2%); Historia, Geografía y Ciencias Sociales [CS] (395 casos, 12%); Inglés [I] (496 casos, 15,1%); Lenguaje y Comunicación [L] (394 casos, 12%); Matemática [M] (495 casos, 12,9%); Química [Q] (119 casos, 3,6%); Religión Católica [RC] (134 casos, 4,1%); y Religión Evangélica [RE] (51 casos, 1,6%).

La comparación de las medias entre las diferentes asignaturas de la muestra EM se presenta en la figura 7.

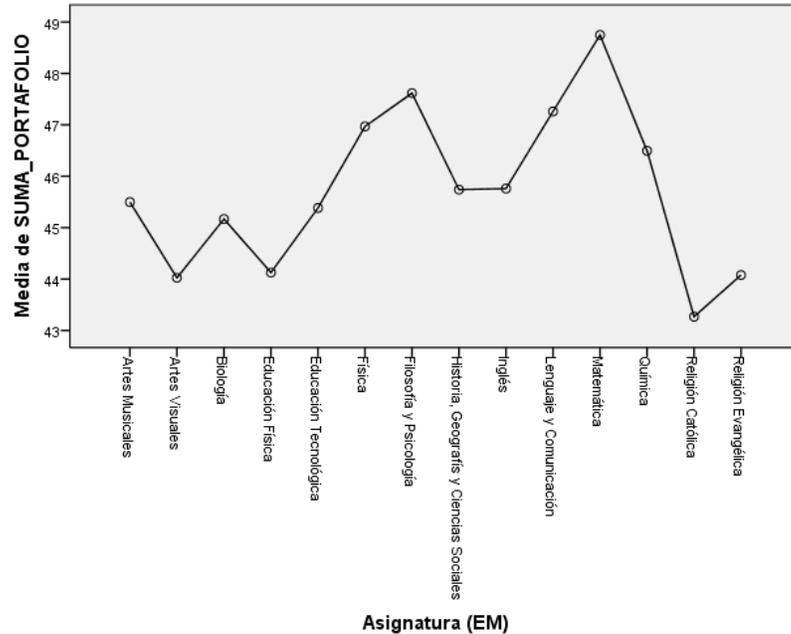


Figura n°7. Comparación de medias en la variable *Asignatura (Educación Media)*

Muestra EE:

En esta muestra se seleccionaron 1.968 casos, distribuidos en 2 grupos: 1=Trastorno Específico del Aprendizaje [TEA] (1444 casos, 73,4%); y 2=Trastorno Específico del Lenguaje [TEL] (524 casos, 26,6%).

Muestra EA:

Finalmente, esta muestra concentra 278 casos evaluados en el nivel Educación de Adultos, los que se encuentran divididos en 11 asignaturas: Básica Ciencias Naturales [EAEBCN] (4 casos, 1,4%); Básica Estudios Sociales [EAEBCS] (10 casos, 3,6%);

Básica Lengua Castellana y Comunicación [EAEBL] (18 casos, 6,5%); Básica Educación Matemática [EAEBM] (7 casos, 2,5%); Media Biología [EAEMB] (17 casos, 6,1%); Media Estudios Sociales [EAEMCS] (57 casos, 20,5%); Media Física [EAEMF] (11 casos, 4%); Media Inglés [EAEMI] (45 casos, 16,2%); Media Lengua Castellana y Comunicación [EAEML] (44 casos, 15,8%); Media Educación Matemática [EAEMM] (53 casos, 19,1%); Media Química [EAEMQ] (12 casos, 4,3%).

3.5 Procedimientos de análisis de datos

3.5.1 Análisis descriptivos e inferenciales

En una primera etapa se sometió la base de datos poblacional, así como cada una de las muestras a una serie de análisis descriptivos e inferenciales. Los datos descriptivos de la población y las respectivas muestras se encuentran detallados en las tablas del anexo N°1.

Los análisis inferenciales que se realizaron sobre la base de datos ya tratada, consistieron en pruebas para evaluar la unidimensionalidad y la consistencia interna del instrumento, ambas pruebas se llevaron a cabo sobre el instrumento completo, y sobre módulos 1 y 2 de forma independiente. La evaluación de la unidimensionalidad se realizó aplicando el procedimiento de análisis factorial exploratorio (AFE) por medio de extracción de los componentes principales, en la modalidad de análisis paralelo (PA), en base a la matriz de correlaciones policóricas, dado que este tipo de correlaciones representa un mejor índice cuando se trata de ítems con patrones de respuesta ordinal (Burga, 2006; Domínguez, 2014; Elosua & Zumbo, 2008; Freiberg Hoffmann, Beatriz Stover, De la Iglesia, & Fernández Liporace, 2013; Zumbo, Gadermann, & Zeisser, 2007). De acuerdo con el criterio de Reckase (1979), cuando la varianza explicada por el primer factor es mayor o igual que el 20%, es posible asumir que el test cumple con la

condición de la unidimensionalidad (Dragow & Parsons, 1983; Reckase, 1979). Dado que, como se ha señalado, la puntuación de los indicadores del portafolio responde a un patrón de respuesta politómica ordinal, la consistencia interna fue analizada por medio del índice Alpha ordinal, calculado a partir de las cargas factoriales de cada ítem a partir de la matriz de correlaciones policóricas. Los análisis anteriores se llevaron a cabo utilizando el software FACTOR (Lorenzo-Seva & Ferrando, 2006), y un programa desarrollado específicamente para el cálculo del índice Alpha ordinal por (Domínguez, 2012) este programa fue facilitado vía comunicación directa con el autor.

3.5.2 Análisis de Impacto

Cuando las diferencias en las puntuaciones de un test se deben a diferencias reales en la habilidad o atributo presente en distintos grupos que responden al test, se denomina Impacto (Ackerman, 1992; Camilli & Shepard, 1994), el que se manifiesta, en la medición comparada del atributo entre los grupos, como diferencias en la media, en la desviación estándar o en ambas (Cuevas, 2013). La evaluación para determinar la presencia de Impacto puede llevarse a cabo utilizando una comparación de medias mediante la prueba t de Student, sin embargo, por la sensibilidad de esta prueba frente a muestras de gran tamaño, donde pueden amplificarse pequeñas diferencias y ser consideradas estadísticamente significativas, se recomienda utilizar alguna medida del tamaño del efecto (Fidalgo & Ferreres Traver, 2002).

En virtud de lo anterior, se utilizó, como indicador de impacto, el estadístico d de Cohen (Cohen, 1988), estimando la varianza conjunta de los grupos. Se consideró un impacto pequeño, un valor absoluto de la d mayor que 0.2 y menor a 0.5; un impacto medio, una magnitud de la d mayor o igual a 0.5 y menor a 0.8; y un impacto grande, una magnitud de la d mayor a 0.8 (Cohen, 1992).

A continuación se presenta la ecuación correspondiente a la d de Cohen.

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}}$$

Donde:

$\bar{X}_1 =$ Media del grupo focal

$\bar{X}_2 =$ Media del grupo de referencia

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

$n_1 =$ Número de casos del grupo focal

$n_2 =$ Número de casos del grupo de referencia

3.5.3 Análisis DIF

En el presente estudio se observaron simultáneamente dos estadísticos para determinar la presencia de DIF: QGMH₂, o *estadístico de Mantel-Haenszel generalizado ordinal* (Landis et al., 1978), y el estadístico SMD/S_i (Zwick & Thayer, 1996; Zwick et al., 1997).

Existe una extensa discusión en torno a ciertos aspectos metodológicos de los procedimientos para determinar la presencia de DIF, entre otros, algunos aspectos relevantes son la definición de la variable de emparejamiento, el refinamiento por medio de aplicación bietápica o iterativa de los procedimientos, y la implementación de análisis sobre varios grupos de forma simultánea y las implicancias de dicha aplicación (Fidalgo & Mellenbergh, 1999; Fidalgo, Mellenbergh, & Muñiz, 2000; S. Kim, Cohen, & Park, 1995; Moses et al., 2013; Zwick, 2012; Zwick, Ye, & Isham, 2013).

No existe una conclusión definitiva en torno a ninguno de los problemas planteados arriba. En cuanto a la definición de la variable de emparejamiento a partir de

las puntuaciones observadas se ha discutido la posible circularidad presente en la lógica de utilizar los puntajes obtenidos en un test, posiblemente sesgado, para determinar si sus ítems presentan potenciales sesgos. Frente a este problema, se han implementado mecanismos para purificar la variable de emparejamiento por medio de análisis iterativos que excluyen sucesivamente aquellos ítems identificados con DIF. Al respecto, Fidalgo y Mellenberg (1999) señalan que la implementación de procedimientos de MH, el costo de aplicar un análisis iterativo no se justifica frente al beneficio de una implementación bietápica, por otra parte, Zwick (2012) indica que un refinamiento de esta variable puede ser favorable cuando el DIF no es compensatorio, es decir, cuando se presenta en una dirección, de lo contrario puede conducir a interpretaciones equivocadas, lo mismo si se trata de tamaños muestrales pequeños (Zwick et al., 2013). En cuanto a los procedimientos sobre ítems de respuesta abierta, no se ha llegado a conclusiones definitivas, sin embargo, se llama la atención en torno a los efectos que el proceso de refinamiento tiene en test con poca cantidad de ítems, pudiendo resultar una fuente de distorsión del análisis (Moses et al., 2013).

En cuanto a la presencia de varios grupos de comparación, el ETS observa la dificultad de satisfacer los criterios relativo a tamaños muestrales cuando se definen múltiples grupos, sin embargo, dado el costo de implementar múltiples análisis pareados, recomienda implementarlos cuando es posible (Zwick, 2012). Al respecto, se ha planteado la necesidad de ajustar el nivel de significación estadística en función del error asociado a realizar múltiples comparaciones, aplicando la corrección de Bonferroni (Ferrerres Traver, González-Romá, & Gómez-Benito, 2000; Fidalgo & Ferreres Traver, 2002), sin embargo, esta recomendación no es compartida por el ETS, desde donde se señala que, en el contexto del DIF, es mayor el costo del llamado Error tipo 2 (falsos negativos), que la posibilidad de identificar incorrectamente ítems con DIF (Zwick, 2012).

2.2.3 Estadísticos Generalizados de Mantel-Haenszel

El procedimiento fue propuesto por Mantel-Haenszel (1959) para el análisis de tablas de contingencia tridimensionales, y adaptado para la detección del DIF por Holland y Thayer (1988). Actualmente, la familia de estadísticos de Mantel y Mantel-Haenszel ofrece una gran variedad de posibilidades para el análisis del funcionamiento diferencial, dependiendo del tipo de DIF que se busca identificar, el tipo de ítems, la escala de puntuación y la cantidad de grupos que se comparan. Fidalgo presenta un esquema con una clasificación de estos estadísticos en función del análisis DIF correspondiente.

A partir de los trabajos originales de Natham Mantel y William Haenszel (1959), se han desarrollado una serie de generalizaciones que permiten la aplicación del análisis sobre ítems politómicos, tales como el test MH generalizado (Mantel & Haenszel, 1959), que permite analizar ítems politómicos cuyas categorías de respuesta correspondan a una variable nominal.

Otra alternativa es el test de Mantel (Mantel, 1963), que permite el análisis de repuestas ordenadas entre dos grupos de comparación. Fue el año 1978 cuando Landis et al. propusieron un nuevo estadístico que integraba las generalizaciones anteriores, permitiendo el análisis de ítems politómicos, nominales u ordinales, para múltiples grupos (Fidalgo & Bartram, 2010; Fidalgo & Madeira, 2008; Fidalgo, Quintanilla, Fernández, Pons, & Aguerri, 2010).

Actualmente, la familia de estadísticos de Mantel y Mantel-Haenszel ofrece una gran variedad de posibilidades para el análisis del funcionamiento diferencial, dependiendo del tipo de DIF que se busca identificar, el tipo de ítems, la escala de puntuación y la cantidad de grupos que se comparan.

Fidalgo presenta un esquema con una clasificación de estos estadísticos en función del análisis DIF correspondiente. (Figura nº8)

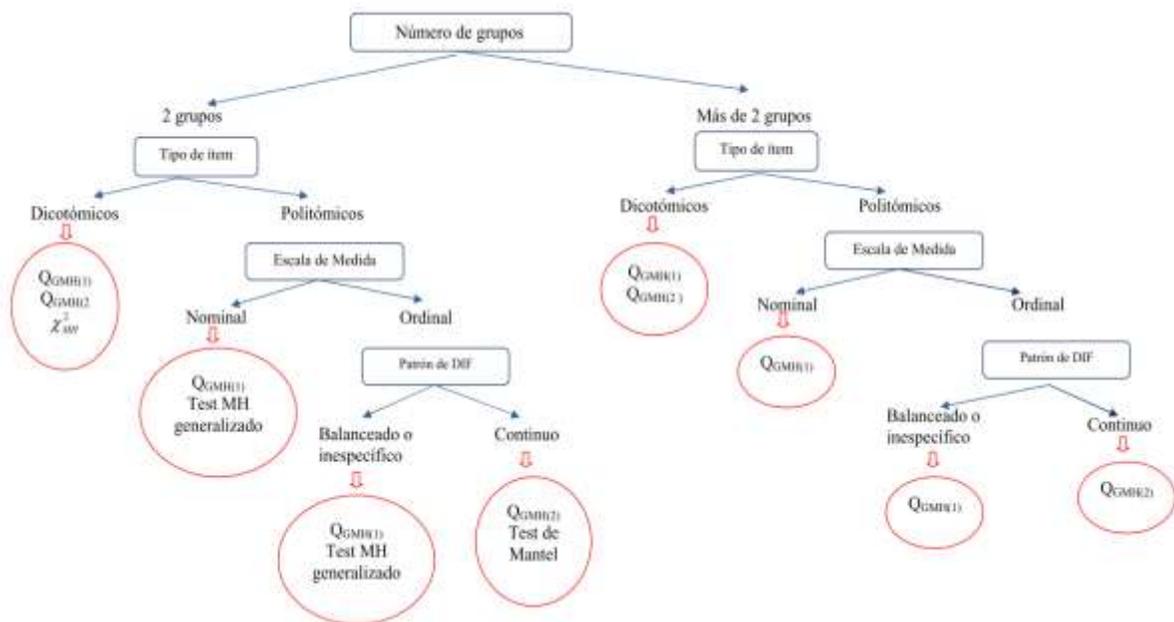


Figura 8. Clasificación de los estadísticos generalizados de MH según el tipo de ítem, escala de medida y patrón del DIF. (Fidalgo, 2010, p.17).

El estadístico Q_{GMH} propuesto por Landis et al. (1978), permite el análisis de Q estratos o tablas de contingencia de dimensiones $R \times C$, siendo R el número de filas y C el número de columnas. En un análisis DIF, cada fila (factor) corresponde a un grupo de comparación, mientras las columnas a las diferentes categorías del ítem; en tanto, los diferentes estratos corresponden a diferentes niveles de la variable de emparejamiento,

esta variable se suele definir a partir de la puntuación total del test, y representa diferentes niveles de habilidad de los evaluados (Ver tabla 1) (Fidalgo & Bartram, 2010; Fidalgo & Mellenbergh, 1999; Fidalgo & Scalón, 2010; S.-H. Kim et al., 2007).

Tabla 1

Estructura de la tabla de contingencia en el h-ésimo estrato.

Niveles del Factor	Categorías de la variable de respuesta				Total
	1	2	j	C	
1	n_{h11}	n_{h12}	n_{h1j}	n_{h1C}	$N_{h1\cdot}$
2	n_{h21}	n_{h22}	n_{h2j}	n_{h2C}	$N_{h2\cdot}$
·	·	·	·	·	·
i	n_{hi1}	n_{hi2}	n_{hij}	n_{hiC}	$N_{hi\cdot}$
·	·	·	·	·	·
R	n_{hR1}	n_{hR2}	n_{hRj}	n_{hRC}	$N_{hR\cdot}$
Total	$n_{h\cdot 1}$	$n_{h\cdot 2}$	$n_{h\cdot j}$	$n_{h\cdot C}$	$N_{h\cdot\cdot}$

Nota: Tabla tomada de Fidalgo (2010)

El estadístico Q_{MH} contrasta la H_0 de no asociación entre el factor y la variable de respuesta, controlando el efecto de la variable de emparejamiento y se define con la siguiente fórmula (Fidalgo, 2011a; Fidalgo & Bartram, 2010; Fidalgo et al., 2010; Fidalgo & Scalón, 2010; Landis et al., 1978).

$$Q_{GMH} = \left\{ \sum_{h=1}^q (\mathbf{n}_h - \mathbf{m}_h)' \mathbf{A}_h' \right\} \left\{ \sum_{h=1}^q \mathbf{A}_h \mathbf{V}_h \mathbf{A}_h' \right\}^{-1} \left\{ \sum_{h=1}^q \mathbf{A}_h (\mathbf{n}_h - \mathbf{m}_h) \right\}$$

Donde:

$$\mathbf{n}_h = (n_{h11}, n_{h21}, \dots, n_{hRC})' \quad (CR \times 1)$$

$$\mathbf{m}_h = N_{h\cdot} (\mathbf{p}_{h\cdot} \otimes \mathbf{p}_{h\cdot}) \quad (CR \times 1)$$

$$\mathbf{A}_h = \mathbf{C}_h \otimes \mathbf{R}_h$$

$$\mathbf{V}_h = N_{h\cdot}^2 / (N_{h\cdot} - 1) \{ (\mathbf{D}_{p_{h\cdot}} - \mathbf{p}_{h\cdot} \mathbf{p}_{h\cdot}') \otimes (\mathbf{D}_{p_{h\cdot}} - \mathbf{p}_{h\cdot} \mathbf{p}_{h\cdot}') \} \quad (CR \times CR)$$

Donde \mathbf{n}_h corresponde a las frecuencias observadas en la medición, \mathbf{m}_h es el vector de frecuencias esperadas, \mathbf{A}_h una matriz de funciones lineales y \mathbf{V}_h la matriz de covarianzas.

Además, $\mathbf{p}_{h\cdot}$ es un vector con las proporciones marginales de las categorías de respuesta ($p_{h\cdot j} = N_{h\cdot j} / N_{h\cdot}$); y $\mathbf{p}_{\cdot h}$ es un vector con las proporciones marginales de los diferentes grupos (factores) ($p_{hi\cdot} = N_{hi\cdot} / N_{h\cdot}$). $\mathbf{D}_{p_{h\cdot}}$ y $\mathbf{D}_{p_{\cdot h}}$ son matrices diagonales, en cuyas diagonales principales se encuentran los elementos de $\mathbf{p}_{h\cdot}$ y $\mathbf{p}_{\cdot h}$, respectivamente.

La versatilidad de este estadístico está dada por la matriz \mathbf{A}_h , ya que sus componentes pueden definirse de acuerdo con la H_1 , según la naturaleza de las categorías de respuesta al ítem. De este modo, para ítems politómicos ordinales la H_1 señala que la media de las respuestas al ítem varía a través de los diferentes niveles del factor, y se utilizará el estadístico Q_{GMH2} , o estadístico generalizado MH ordinal. En este caso, la matriz de linealidad ($\mathbf{A}_h = \mathbf{C}_h \otimes \mathbf{R}_h$), se define por el vector $\mathbf{C}_h = (\mathbf{C}_{h1}, \dots, \mathbf{C}_{hc})$, con dimensiones $(1 \times C)$, donde c_{hj} refleja la naturaleza ordinal de la variable de respuesta en el nivel h -ésimo de la variable de emparejamiento, y $\mathbf{R}_h = [\mathbf{I}_{R-1}, -\mathbf{J}_{R-1}]$, donde, \mathbf{I}_{R-1} es una matriz de identidad de dimensión $(R-1 \times R-1)$, y \mathbf{J}_{R-1} es un vector de unos.

Bajo la hipótesis nula de no asociación, *el estadístico* $Q_{GMH(2)}$ sigue aproximadamente una distribución ji-cuadrado con $gl = (R-1)$. (Fidalgo, 2005, 2011a, 2011b; Fidalgo & Bartram, 2010; Fidalgo & Madeira, 2008; Fidalgo & Scalón, 2010; Gómez-Benito & Hidalgo, 2003; Gómez-Benito et al., 2005).

Los análisis pareados se realizaron en una aplicación bietápica, excluyendo de la variable de emparejamiento los ítems identificados con DIF en la primera etapa. Este tipo de análisis se implementó para todas las variables.

Los análisis multigrupo se implementaron para las variables *Edad* y *Nivel de Enseñanza*. Este procedimiento, en una primera etapa, compara todos los grupos simultáneamente, luego se purifica la variable de emparejamiento excluyendo los ítems que mostraron asociaciones significativas, y luego se vuelve a realizar el análisis global, a continuación los ítems identificados con DIF en la segunda etapa, son sometidos a comparaciones pareadas con el nivel de significación ajustado por medio de la aplicación de la corrección de Bonferroni (nivel de significación dividido por la cantidad de comparaciones)

En la variable *Asignatura* no fue posible implementar un análisis multigrupo para obtener el estadístico Q_{GMH2} , ya que al concluir la primera etapa del análisis los valores resultaban significativos para todos los ítems, al nivel de significancia $\alpha=0.05$, lo que impidió al software GMHDIF calcular la variable de emparejamiento, deteniendo el análisis antes de concluir. Se intentaron pruebas ajustando el nivel de significancia por medio de la corrección de Bonferroni, sin embargo el resultado fue el mismo.

Finalmente, se aplicó un análisis por parejas de grupos (GF-GR), ajustando el nivel de significación (α) por medio de la corrección de Bonferroni (α/N° de comparaciones pareadas). En este nuevo análisis solamente se incluyeron los ítems que presentaron DIF en el análisis con el nivel de significación $\alpha= 0.05$. El valor corregido para la significación de la prueba fue de $\alpha= 0.00556$ para las 9 comparaciones pareadas correspondientes a la variable *Asignatura* (segundo ciclo), donde fue excluido el grupo Ciencias naturales por no presentar diferencias significativas con respecto al grupo de referencia en las pruebas inferenciales.

2.2.4 Diferencia de Medias Estandarizadas

El método de estandarización, aplicado a ítems politómicos de respuesta ordenada, corresponde a una extensión de la formulación de Dorans y Scmitt (1991), (Zwick & Thayer, 1996; Zwick et al., 1997). En este procedimiento las medias de los grupos de referencia y focal se ponderan por medio de un factor que equipara la distribución de ambos grupos, a partir de la distribución del grupo focal. El estadístico original, corresponde a la diferencia entre la media de grupos focal y referencia, estandarizada por medio del factor de ponderación (SMD). Ver Tabla 2

Tabla 2
Representación de los ítems politómicos

Grupo	Categoría de respuesta					Total
	y ₁	y ₂	y ₃	...	y _M	
Referencia	n _{R1k}	n _{R2k}	n _{R3k}	...	n _{RMk}	n _{R+k}
Focal	n _{F1k}	n _{F2k}	n _{F3k}	...	n _{FMk}	n _{F+k}
Total	n _{+1k}	n _{+2k}	n _{+3k}	...	n _{+Mk}	N _{++k}

(Elosua, 2002, 2006)

$$SMD = \sum_k \frac{n_{F+k}}{n_{F++}} \frac{\sum_m y_m n_{Fmk}}{n_{F+k}} - \sum_k \frac{n_{R+k}}{n_{R++}} \frac{\sum_m y_m n_{Rmk}}{n_{R+k}}$$

El índice SMD se puede normalizar para que su medida no dependa de la escala de respuesta al ítem, para esto Zwick (1996) propone dividir el índice SMD por la desviación estándar obtenida en el ítem, combinando los grupos de referencia y focal. El nuevo estadístico es SMD/σ_i .

Donde:

$$\sigma_i = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

Los análisis del estadístico Q_{GMH2} se llevaron a cabo utilizando el software GMHDIF (Fidalgo, 2011a, 2011b). El (*Educational Testing Service* (ETS) utiliza un criterio para clasificar el DIF en 3 categorías, A, B y C, insignificante, moderado y severo, respectivamente. Para el presente trabajo se adoptará el criterio propuesto por Zwick (1996, 1997), de acuerdo con el cual, se evaluará la presencia de DIF observando conjuntamente la significación del estadístico utilizado ($\alpha= 0,05$), y la correspondencia a las categorías de clasificación dada por la magnitud del estadístico SMD/σ_i , que según la propuesta de Zwick (1996, 1997), señala que un valor absoluto menor a 0.17 será indicador de un DIF insignificante, una magnitud igual o mayor que 0.17 y menor o igual que 0.25 corresponde a un DIF moderado, mientras un DIF severo corresponderá a una magnitud mayor a 0.25.(Elosua, 2006; Elosua & López-Jáuregui, 2007; Zwick & Thayer, 1996; Zwick et al., 1997).

IV RESULTADOS

En el presente apartado se expondrán, de forma extensa, los resultados obtenidos en cada uno de los análisis implementados. En primer lugar, se expondrán los resultados correspondientes a análisis inferenciales que evalúan, tanto la unidimensionalidad del portafolio, como la consistencia interna de este instrumento y sus respectivos módulos.

Posteriormente, de acuerdo con los objetivos específicos definidos para el estudio, se reportan los resultados correspondientes a los análisis de *Impacto*, y los diferentes procedimientos de análisis DIF implementados para cada una de las variables: *Sexo, Zona Geográfica, Edad, Nivel de Enseñanza y Asignatura*, respectivamente. En cada caso se presenta una tabla que resume los resultados, según el criterio de convergencia de los resultados, es decir, se presentan únicamente aquellos indicadores cuyos resultados señalan la presencia de DIF simultáneamente, tanto para los análisis MH, como para las pruebas basadas en estandarización.

Finalmente, para simplificar la lectura de los resultados y facilitar la interpretación de los mismos, se presenta un cuadro resumen, en el que se incluyen todos los indicadores que presentan DIF severo en alguna de las comparaciones analizadas.

4.1 Análisis Inferenciales:

El AFE en modalidad de Análisis Paralelo (PA), en base a la matriz de correlaciones policóricas, extracción de componentes principales, arrojó que el primer factor, explica un 19,5% de la varianza total del instrumento, a partir de este resultado no es posible determinar que el portafolio cumpla con el supuesto de unidimensionalidad según el criterio de Reckase (1979). En cuanto a su consistencia interna, esta fue evaluada calculando el coeficiente alpha ordinal a partir de los pesos factoriales

entregados por la matriz de correlaciones policóricas ($\alpha = .823$). El nivel de consistencia mostrado se considera aceptable (Elosua & Zumbo, 2008; Zumbo et al., 2007)

A pesar de que la extracción de componentes principales muestra que el primer factor explica un porcentaje de varianza cercano al límite que define el criterio de Reckase (1979), no es posible asumir el supuesto de la unidimensionalidad para el instrumento completo. Posteriormente, se procedió a evaluar si este principio se cumple para cada uno de los módulos del portafolio. Los análisis se realizaron replicando las características ya descritas, obteniéndose para ambos Módulos un coeficiente alpha ordinal = .70. Se aprecia la disminución del coeficiente alpha ordinal en relación con el obtenido para el test completo, esto puede explicarse por la sensibilidad de este índice a la disminución de los ítems incorporados en el análisis. Por otra parte, la matriz de correlaciones policóricas, arrojó un primer factor que explica el 29% de la varianza total para el Módulo 1, y un 33,51% para el Módulo 2.

Los índices de consistencia interna se consideran aceptables, y la proporción de varianza explicada permite aceptar el principio de unidimensionalidad de cada Módulo.

4.2 Resultados por variables

4.2.1 Sexo

Evaluación de Impacto

El análisis de impacto realizado comparando las medias de los grupos a través del estadístico d de Cohen, solamente mostró niveles de impacto nulo o pequeño para cada uno de los indicadores del portafolio, siendo el ítem 6.1 el que presentó una mayor diferencia en favor del grupo de referencia (femenino), con una $d = -3.7$, valor considerado pequeño.

Análisis DIF

La aplicación del Test de Mantel con la muestra pequeña ($N = 1000$) arrojó 6 ítems que presentan DIF, correspondientes a los indicadores 1.1 (*Formulación y coherencia entre los objetivos*); 2.1 (*Características de los alumnos consideradas en la unidad*); 6.1 (*Calidad del inicio de la clase*); 7.1 (*Calidad de las explicaciones*); 7.3 (*Calidad de la retroalimentación a los alumnos*) y 7.4 (*Indicador por asignatura o área*). Para calcular la variable de emparejamiento se omitieron 4 ítems que en la primera etapa del análisis presentaron asociación sobre el nivel de significancia (0.05), estos ítems corresponden a los indicadores 1.1; 6.1; 7.1 y 7.4.

El análisis aplicado sobre la muestra grande ($N=10000$) mostró asociación en 14 de los ítems, esto implicaría que solamente 6 ítems analizados, correspondientes a los indicadores 2.2 (*Análisis de la unidad realizada*); 3.2 (*Adecuación de las evaluaciones a los objetivos de aprendizaje de la unidad realizada*); 4.1 (*Atribución de los resultados de aprendizaje*); 4.2 (*Calidad de la retroalimentación*); 6,3 (*Contribución de las actividades al logro de los objetivos de la clase*) y 7,2 (*Calidad de las preguntas*) no presentarían DIF. Es necesario observar que en la primera etapa del procedimiento se excluyeron 18 de los 20 ítems analizados, calculando la variable de emparejamiento

únicamente con los ítems 1.3 (*Calidad de la secuencia entre las clases de la unidad*) y 3.1 (*Calidad de la Evaluación*).

El método de estandarización arrojó que el ítem 6.1 presenta DIF severo en contra del grupo focal ($SMD/\sigma_i = -0.28$), mientras el ítem 7.4 presenta DIF moderado ($SMD/\sigma_i = 0.25$) en contra del grupo de referencia.

Al aplicar el criterio de convergencia de métodos, los indicadores 6.1 y 7.4 presentan DIF, en el primer caso el ítem tendría funcionamiento diferencia contra el grupo focal, de magnitud severa, y en el segundo caso corresponde, a una magnitud moderada, contra el grupo de referencia.

La tabla N°3 muestra los ítems que presentan DIF al considerar la convergencia de los análisis generalizados de Mantel Haenszel, con dos tamaños muestrales y el método de estandarización, cuyo estadístico SMD/σ_i determina, por medio de su valor absoluto, la clasificación del DIF. Los resultados detallados de cada análisis se encuentran disponibles en el acápite *anexos*, del presente trabajo

Tabla 3
Análisis Impacto y DIF variable Sexo

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	Análisis GMH N= 1000		Análisis GMH N= 10000		SMD	
		<i>d</i>	Q _{MH2}	<i>p</i>	Q _{MH2}		
6.1) Calidad del inicio de la clase	-0.37	6.26	0.01	99.54	0.00	-0.28	Severo
7.4) Indicador por asignatura o área	0.14	7.44	0.01	102.57	0.00	0.23	Moderado*

*DIF contra el grupo de referencia. Grupo de Referencia (GR)=Femenino v/s Grupo Focal (GF)=Masculino

4.2.2 Zona Geográfica

Evaluación de Impacto

La evaluación de impacto mostró que no existen magnitudes medianas o grandes de la d de Cohen, siendo el indicador 7.1 el que presentó el valor absoluto más elevado (d -Cohen = -1.1).

Análisis DIF

Test de Mantel (N=1000) mostró 6 ítems con DIF, correspondientes a los indicadores 1.2, 2.2, 7.1, 7.2, 7.3 y 7.4. La misma prueba para una muestra de 10000 casos arrojó DIF para los indicadores 1.2, 2.2, 3.1, 5.1, 5.3, 7.1, 7.3 y 7.4. La variable de emparejamiento se calculó sobre los ítems que no mostraron DIF en la primera etapa del análisis, por lo que se excluyeron los mismos ítems mencionados además del ítem 6.1.

El estadístico SMD/ σ_i mostró asociación moderada en contra del grupo focal en los ítems 2.2, 4.2, 7.1 y 7.2.

La tabla N°4 muestra los ítems que presentan DIF al aplicar el criterio de convergencia de los análisis aplicados:

Tabla 4
Análisis Impacto y DIF variable Zona Geográfica

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	Análisis GMH N= 1000		Análisis GMH N= 10000		SMD	
		d	Q_{MH2}	p	Q_{MH2}	p	
2.2) Análisis de la unidad realizada	-0.08	9.37	0.00	7.50	0.01	-0.17	Moderado
7.1) Calidad de las explicaciones	-0.11	7.84	0.01	14.26	14.26	-0.20	Moderado

Nota: GR=Urbana v/s
GF=Rural

4.2.3 Edad

Evaluación de Impacto

La evaluación de impacto únicamente mostró una magnitud media ($d = -0.54$) en el indicador 6.1, en contra del grupo focal, para la comparación pareada entre los grupos G3 (31 a 35 años) y G8 (61 años o más), correspondientes a grupos de referencia y focal respectivamente.

Análisis DIF

Fueron aplicados dos procedimientos de Mantel Haenszel para la identificación de posibles ítems con DIF. El primero consistió en el análisis pareado de cada uno de los grupos que presentó diferencias significativas con respecto al grupo de referencia en las pruebas inferenciales, y el segundo correspondió a un análisis multigrupo, donde se contrastaron simultáneamente las asociaciones entre todos los grupos que presentaron diferencias significativas con respecto al grupo de referencia.

A partir de análisis pareado, no se detectaron ítems con DIF para la comparación de los grupos G3 (36 a 40 años) y G4 (41 a 45 años).

La comparación de los grupos G3 y G6 (51 a 55 años) arrojó que los indicadores 5.3 y 7.1 presentan DIF, para una significación $\alpha=0.05$.

El análisis pareado de los grupos G3 y G7 (56 a 60 años) muestra 4 indicadores con DIF (1.1, 2.2, 5.3 y 7.4).

Entre los grupos G3 y G8 (60 años o más) se presentaron 7 indicadores con DIF para el valor $\alpha=0.05$ (1.1, 2.2, 4.2, 5.3, 6.1, 6.2 y 7.4).

El análisis multigrupo mostró, en su primera fase, que los indicadores 1.1, 2.2, 5.3, 6.1, 6.2 y 7.4 presentaron asociaciones significativas para el nivel $\alpha=0.05$. La segunda fase del análisis sometió a evaluación únicamente los indicadores que, en el primer análisis se rechazó la H_0 de no asociación para el valor $\alpha=0.05$, realizando análisis pareados y ajustando el nivel de significación por medio de la corrección de Bonferroni (α / N° de comparaciones pareadas). Para el nivel de significación ajustado

($\alpha=0.0125$), el indicador 1.1 presentó DIF para la comparación de los grupos G3/G7 y G3/G8; asimismo, el indicador 2.2 mostró DIF para los grupos G3/G7. El indicador 5.3 presentó DIF en las comparaciones de los grupos G3/G6, G3/G7 y G3/G8; el indicador 6.1 presentó DIF para la prueba entre los grupos G3/G8; y el indicador 7.4 mostró presencia de DIF al comparar los grupos G3/G8.

El procedimiento basado en el método de estandarización, mostró un DIF moderado en el indicador 5.2, en contra del grupo de referencia en las comparaciones de los grupos G3/G4 y G3/G6, y DIF severo contra el grupo de referencia en las comparaciones G3/G7 y G3/G8.

Del mismo modo, la comparación entre los grupos G3/G8 también mostró DIF moderado contra el grupo de referencia en el indicador 7.4. Por otra parte, en la comparación de estos grupos los indicadores 2.2 y 5.3 presentaron DIF moderado y el indicador 6.1 DIF severo en contra del grupo focal.

La tabla n°5 muestra el resumen de los indicadores que han sido identificados con DIF al aplicar el criterio de convergencia de los procedimientos de análisis.

Tabla 5

Indicadores con DIF. Convergencia de procedimientos, variable Edad

	GMH						SMD	Clasificación del DIF
	Análisis Pareado		Análisis Multigrupo				SMD/ σ_i	
	G3 vs G8		global		pareado			
	Q_{MH2}	$\alpha=0.05$	Q_{MH2}	$\alpha=0.05$	Q_{MH2}	$\alpha=0.0125$		
1.1) Formulación y coherencia entre los objetivos	8.68	0.00	17.01	0.00	11.98	0.00	-0.19	Moderado
5.3) Supervisión de las actividades	8.77	0.00	16.09	0.00	7.88	0.01	-0.19	Moderado
6.1) Calidad del inicio de la clase	7.88	0.01	11.35	0.02	7.15	0.01	-0.30	Severo
7.4) Indicador por asignatura o área	13.70	0.00	13.58	0.01	6.51	0.01	0.21	Moderado*

*DIF contra el grupo de referencia

4.2.4 Nivel de Enseñanza

4.2.4.1 Primer Ciclo como GR

Evaluación de Impacto

El estadístico d de Cohen mostró magnitudes irrelevantes de impacto en la mayoría de los indicadores, en las cuatro comparaciones pareadas implementadas. El indicador 1.2 presentó un tamaño de impacto medio en contra de los grupos focales Enseñanza Media y Educación de Adultos ($d=-0.50$ y $d=-0.53$, respectivamente). El indicador 1.3 presentó impacto medio contra el grupo focal Enseñanza Media ($d=-0.51$). También se observó impacto de magnitud media contra el grupo focal Educación Parvularia, en los indicadores 6.6 ($d=-0.61$) y 7.4 ($d=-0.62$) e impacto de magnitud grande en el indicador 7.3 ($d=-0.85$).

Análisis DIF

El análisis pareado para calcular el estadístico Generalizado Mantel Haenszel muestra asociaciones significativas, al nivel $\alpha=0.05$, para los grupos Primer Ciclo Básico (PC) y Educación Parvularia (EP), en los indicadores 1.2, 1.3, 2.2, 3.1, 3.2, 4.14.2, 5.1, 5.3, 6.1, 6.2, 7.2, 7.3 y 7.4.

Al comparar los grupos PC y Segundo Ciclo Básico (SC) mostraron DIF los siguientes indicadores: 1.1, 1.2, 1.3, 2.2, 2.3, 3.1, 3.2, 4.1, 5.1, 5.3, 6.1, 7.1, 7.2 y 7.3.

El análisis pareado de los grupos PC y Educación Media (EM) muestra presencia de DIF en los indicadores 1.1, 1.2, 2.1, 2.2, 2.3, 3.1, 3.2, 4.2, 5.1, 5.2, 6.1, 7.1, 7.2, 7.3 y 7.4. El software GMHDIF no pudo calcular el estadístico QMH2 para el indicador 4.1 en el análisis de estos dos grupos.

Finalmente, la pareja de grupos PC y Educación de Adultos (EA) presentó DIF para los indicadores 1.2, 1.3, 2.1, 5.1, 5.2, 6.1, 7.1 y 7.2.

El procedimiento multigrupo entregó como resultado de la primera fase del análisis, asociación para el nivel $\alpha=0.05$ en los indicadores 1.2, 1.3, 2.1, 3.1, 3.2, 5.3, 6.1, 6.2, 7.1, 7.2, 7.3 y 7.4. La segunda fase del análisis, aplicada únicamente sobre los indicadores que presentaron DIF en la prueba global, utilizó un ajuste del nivel de significación $\alpha=0.0125$, luego de aplicar la corrección de Bonferroni. Como resultado de esta prueba se observa que el indicador 1.2 presenta DIF en las comparaciones con los grupos SC y EM, el indicador 1.3 muestra DIF en el contraste con SC, mientras el indicador 2.1 presenta DIF en la comparación con el grupo EA.

Los indicadores 3.1 y 3.2 muestran funcionamiento diferencial en la comparación del grupo de referencia con los grupos SC y EM. Se observa la presencia de DIF en el indicador 5.3 para el contraste con el grupo EP, también para los indicadores 6.1 y 6.2 en las comparaciones con EM y EP respectivamente.

Finalmente, el indicador 7.1 presenta DIF para la prueba con el grupo EM, el indicador 7.2 muestra DIF en el contraste con los grupos EP y SC, y los indicadores 7.3 y 7.4 funcionan diferencialmente en la comparación del grupo PC con EP.

El procedimiento basado en la estandarización mostró, para el análisis pareado de los grupos PC y EP, que los indicadores 1.3, 6.2, 7.3 y 7.4 presentan magnitudes para el SMD/σ_i que indican la presencia de DIF severo en contra del grupo Focal (Educación Parvularia). Al mismo tiempo, el indicador 4.1 presenta DIF moderado, y los indicadores 4.2, 5.1, 5.2, 5.3, 6.1 y 7.2 muestran un SMD/σ_i con valores positivos y una magnitud que muestra DIF severo en contra del grupo del grupo de referencia.

El criterio de convergencia de métodos aplicado a esta comparación pareada de grupos los grupos Primer Ciclo Básico y Educación parvularia, arroja presencia de DIF para 5 indicadores, como se muestra en la Tabla n°6:

Tabla 6
Convergencia de Análisis DIF variable Nivel de Enseñanza

	GMH						SMD	Clasificación del DIF
	Análisis Pareado			Análisis Multigrupo				
	PC-EP		global	pareado		pareado		
	QMH2	$\alpha=0.05$		QMH2	$\alpha=0.0125$			
	QMH2	$\alpha=0.05$	QMH2	$\alpha=0.0125$	SMD/ σ_i			
5.3) Supervisión de las actividades	40.04	0.00	19.477	0.00	10.96	0.00	0.3	Severo*
6.2) Calidad del cierre de la clase	78.12	0.00	16.057	0.00	8.88	0.00	-0.37	Severo
7.2) Calidad de las preguntas	126.52	0.00	18.362	0.00	10.01	0.00	0.58	Severo
7.3) Calidad de la retroalimentación a los alumnos	190.2	0.00	38.637	0.00	26.11	0.00	-0.59	Severo
7.4) Indicador por asignatura o área	92.53	0.00	15.286	0.00	9.84	0.00	-0.38	Severo*

*DIF contra el grupo de referencia. GR=Primer ciclo Básico (PC) vs GF= Educación Parvularia (EP)

Los resultados de la aplicación del procedimiento de estandarización en el análisis pareado de los grupos PC y Segundo Ciclo Básico, muestran 2 indicadores con DIF fuerte (1.2 y 3.1) y un indicador con DIF moderado (3.2) en contra del grupo focal. También se observan 2 indicadores con DIF severo (5.2 y 7.2) y 2 con DIF moderado (7.1 y 7.3) contra el grupo de referencia.

La convergencia del estadístico Generalizado de MH obtenido por los procedimientos de análisis pareados y multigrupo, además del estadístico SMD/σ_i muestra 4 indicadores con DIF.

La comparación entre los grupos Primer Ciclo Básico y Educación Media por el método de estandarización muestra la presencia de DIF severo contra el grupo focal (EM) en los indicadores 1.2, 3.1, 3.2, y 6.1, del mismo modo, se observa DIF moderado en el indicador 5.3. También los resultados muestran valores positivos de SMD/σ_i , esto

significa presencia de DIF en contra del grupo focal en los indicadores, tanto en el indicador 2.1, de magnitud moderada, y DIF severo en los indicadores 5.2, 7.1, 7.2 y 7.4 (Tabla 7).

Tabla 7
Convergencia de Análisis DIF variable Nivel de Enseñanza

Indicadores	GMH						SMD	Clasificación del DIF
	Análisis Pareado		Análisis Multigrupo					
	PC-SC		global		pareado			
	QMH2	$\alpha=0.05$	QMH2	$\alpha=0.05$	QMH2	$\alpha=0.0125$	SMD/ σ_i	
1.2) Relación entre actividades y objetivos	165.85	0.00	18.32	0.00	10.37	0.00	-0.34	Severo
3.1) Calidad de la evaluación	150.62	0.00	35.51	0.00	27.28	0.00	-0.31	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	51.91	0.00	16.33	0.00	7.65	0.01	-0.18	Moderado
7.2) Calidad de las preguntas	196.31	0.00	18.36	0.00	16.32	0.00	0.37	Severo*

*DIF contra el grupo de referencia. GR=Primer ciclo Básico (PC) vs GF= Segundo Ciclo Básico (SC)

La tabla 8 presenta la convergencia de los resultados obtenidos a partir de la aplicación de los métodos de análisis implementados. En resumen, para el par de grupos Primer Ciclo Básico, y Educación Media, se observan 5 indicadores con DIF para esta comparación, todos de magnitud severa, en el caso del indicador 7.1, contra el grupo de referencia.

Tabla 8
Convergencia de Análisis DIF variable Nivel de Enseñanza. Primer Ciclo Básico vs Educación Media

Indicadores	GMH						SMD	Clasificación del DIF
	Análisis Pareado		Análisis Multigrupo					
	PC-EM		global		pareado			
	QMH2	$\alpha=0.05$	QMH2	$\alpha=0.05$	QMH2	$\alpha=0.0125$	SMD/ σ_i	
1.2) Relación entre actividades y objetivos	67.31	0.00	18.32	0.00	0.294	0.01	-0.38	Severo
3.1) Calidad de la evaluación	77.01	0.00	35.51	0.00	0.929	0.00	-0.4	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	40.99	0.00	16.33	0.00	0.001	0.00	-0.28	Severo
6.1) Calidad del inicio de la clase	27.80	0.00	30.60	0.00	0.07	0.00	-0.26	Severo
7.1) Calidad de las explicaciones	127.66	0.00	13.52	0.01	2.082	0.00	0.36	Severo*

*DIF contra el grupo de referencia. GR=Primer ciclo Básico (PC) vs GF= Educación Media (EM)

El último análisis de medias estandarizadas para el grupo de referencia Primer Ciclo Básico se realizó en relación al grupo Educación de Adultos. Los indicadores 1.2, 1.3 y 5.2 mostraron un valor para el estadístico SMD/σ_i con magnitud que da cuenta de un DIF severo contra el grupo focal, asimismo, los indicadores 3.2, 5.3 y 6.2 mostraron la presencia de DIF moderado. Por otra parte, los indicadores 2.2 y 7.2 presentaron DIF en contra del grupo de referencia, de magnitud severa.

Al aplicar el criterio de convergencia de los métodos aplicados, es decir test Generalizado MH aplicado en análisis pareados, test Generalizado MH multigrupo, y Diferencia de Media Estandarizadas (normalizada por la desviación estándar de los grupos combinados), no se encontraron indicadores que presenten DIF de forma simultánea en los análisis implementados para la comparación de los grupos Primer Ciclo Básico y Educación de Adultos, en la variable Nivel de Enseñanza.

4.2.4.2 Educación Especial como GR

Evaluación de Impacto

El indicador 1.2 presentó impacto de magnitud media contra todos los grupos focales: Educación Parvularia ($d=-0.59$), Segundo Ciclo Básico ($d=-0.68$), Enseñanza Media ($d=-0.68$) y Educación de Adultos ($d=-0.69$).

El indicador 6.1 presentó impacto medio contra el grupo focal Educación de Adultos ($d=-0.54$), en esta misma dimensión, el indicador 6.2 muestra impacto de magnitud grande ($d=-0.82$) contra el grupo Educación Parvularia y el indicador 6.3 presenta impacto medio contra los grupos Educación Parvularia ($d=-0.55$) y Educación de Adultos ($d=-0.59$).

El indicador 7.3 mostró una magnitud media, aunque en el límite de esta clasificación, para el estadístico $d=-0.79$, mostrando la presencia de impacto contra el

grupo Educación Parvularia. Finalmente, el indicador 7.4 muestra impacto a favor de tres grupos focales (o contra el grupo de referencia): Segundo Ciclo Básico ($d=0.68$), Educación Media ($d=0.88$) y Educación de Adultos ($d=0.50$).

Análisis DIF

El análisis de las comparaciones pareadas del grupo de referencia Educación especial, mostró presencia de DIF en los indicadores 2.1, 2.3, 3.2, 4.2, 5.2, 6.3 y 7.4 frente al grupo de referencia Educación Parvularia. Frente al grupo focal Segundo Ciclo Básico, se observó que 18 de los 20 indicadores mostraban DIF, exceptuando únicamente los indicadores 2.3 y 4.2.

La comparación pareada contra el grupo focal Educación Media arrojó presencia de DIF en 19 indicadores, al mismo tiempo que no fue posible calcular el estadístico generalizado MH para el indicador 4.1. Por último, la comparación frente al grupo Educación de Adultos mostró DIF para los indicadores 1.1, 1.2, 2.1, 2.2, 4.1, 5.1, 5.2, 5.3, 6.1, 6.2, 6.3, 7.1 y 7.4.

El análisis multigrupo implementado para esta variable arrojó en la etapa de análisis global DIF en los indicadores 1.2, 2.1, 2.2, 3.1, 3.2, 4.2, 5.3, 6.1, 6.2, 7.1, 7.2, 7.3 y 7.4. Estos indicadores fueron sometidos a análisis pareados ajustando el nivel de significación según la corrección de Bonferroni, el nivel de significación ajustado fue $\alpha=0.0125$, y el nuevo análisis mostró DIF en los indicadores 4.2, 5.3, 6.2, 7.2, 7.3 y 7.4 frente al grupo focal Educación Parvularia; en los indicadores 1.2, 2.1, 2.2, 3.1, 3.2, 7.1, 7.2, 7.3 y 7.4 frente al grupo Segundo Ciclo Básico; en los indicadores 1.2, 2.1, 3.1, 3.2, 6.1, 6.2, 7.1 y 7.4 en el análisis pareado frente al grupo Educación Media; y DIF en los indicadores 2.1 y 7.4 frente al grupo focal Educación de Adultos.

El procedimiento SMD arrojó presencia de DIF moderado contra el grupo focal Educación Parvularia en el indicador 1.2 y DIF severo contra este mismo grupo en los indicadores 6.2, 6.3 y 7.3. También se encontró la presencia de DIF en favor de este grupo focal en los indicadores 4.1, 6.1 y 7.4 (moderado), y 4.2, 5.2 y 7.2 (severo).

El resumen de los indicadores que presentan DIF para la comparación de los grupos Educación especial y Educación Parvularia se observa en la Tabla 9.

Tabla 9
Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación Parvularia

Indicadores	IMPACTO		DIF						Clasificación del DIF
	D de Cohen	GMH Pareado		GMH Multigrupo				SMD	
		Q _{MH2}	α=0.05	Global α=0.05		Pareado α=0.0125			
				Q _{MH2}	α=0.05	Q _{MH2}	α=0.0125		
D	Q _{MH2}	α=0.05	Q _{MH2}	α=0.05	Q _{MH2}	α=0.0125	SMD/σ _i		
4.2) Calidad de la Retroalimentación	0.12	5.87	0.02	11.84	0.02	10.35	0.00	0.39	Severo*
5.2) Promoción de la participación de todos los alumnos	-0.28	6.14	0.01	5.19	0.27			0.38	Severo*
6.1) Calidad del inicio de la clase	0.05	0.97	0.32	23.73	0.00	4.24	0.04	0.23	Moderado*
6.2) Calidad del cierre de la clase	-0.82	0.04	0.85	25.34	0.00	13.75	0.00	-0.62	Severo

*DIF contra el grupo de referencia. GR=Educación Especial (EE) vs EP= Educación Parvularia (EP)

El estadístico SMD/σ_i, para la comparación frente al grupo focal Segundo ciclo Básico, muestra presencia de DIF moderado en los indicadores 3.2, 5.1, 5.3, 6.1 y 6.2, y DIF severo en los indicadores 1.2, 2.2 y 6.3, contra el grupo focal. También se identificó presencia de DIF moderado contra el grupo de referencia en los indicadores 1.1, 1.3, 2.1, 7.1 y 7.2; y DIF severo contra el mismo grupo en los indicadores 5.2, 7.3 y 7.4.

La Tabla 10 muestra la totalidad de indicadores con DIF de acuerdo con el criterio utilizado.

Tabla 10
Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Segundo Ciclo Básico

Indicadores	IMPACTO	DIF							Clasificación del DIF
	D de Cohen	GMH Pareado		GMH Multigrupo				SMD/σ _i	
		Q _{MH2} 5	α=0.0	Global α=0.05		Pareado α=0.0125			
				Q _{MH2} 5	α=0.0	Q _{MH} 2	α=0.0125 5		
1.2) Relación entre actividades y	-0.68	201.19	0.00	44.73	0.00	32.58	0.000	-0.49	Severo
2.1) Características de los alumnos	0.06	23.52	0.00	20.35	0.00	9.44	0.002	0.17	Moderado*
2.2) Análisis de la unidad realizada	-0.47	89.60	0.00	12.46	0.01	11.38	0.001	-0.32	Severo
3.2) Adecuación de la evaluación a los	-0.34	28.06	0.00	23.25	0.00	13.58	0.000	-0.17	Moderado
7.1) Calidad de las explicaciones	0.09	24.80	0.00	16.58	0.00	8.67	0.003	0.19	Moderado*
7.2) Calidad de las preguntas	0.09	21.14	0.00	10.57	0.03	11.20	0.001	0.18	Moderado*
7.3) Calidad de la retroalimentación a	0.13	40.38	0.00	41.65	0.00	11.19	0.001	0.26	Severo*
7.4) Indicador por asignatura o área	0.68	432.38	0.00	102.95	0.00	79.49	0.000	0.76	Severo*

*DIF contra el grupo de referencia. GR=Educación Especial (EE) vs EP= Segundo Ciclo Básico (SC)

La comparación entre los grupos Educación Especial y Educación Media por medio del procedimiento SMD, arrojó como resultado la presencia de DIF moderado en los indicadores 2.2 y 3.1, y DIF severo en los indicadores 1.2, 3.2, 5.1, 5.3, 6.1, 6.2 y 6.3, contra el grupo focal. Por otra parte, el indicador 1.3 presentó DIF moderado y los indicadores 2.1, 7.1 y 7.4 mostraron DIF severo contra el grupo de referencia.

Siete indicadores presentan DIF severo y uno presenta DIF moderado al considerar la convergencia de los análisis implementados para la comparación de estos grupos. Ver Tabla 11.

Tabla 11
Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación Media

Indicadores	IMPACTO	DIF								Clasificación del DIF
	D de Cohen	GMH Pareado		GMH Multigrupo				SMD	SMD/ σ_i	
		Q _{MH2}	$\alpha=0.05$	Global $\alpha=0.05$		Pareado $\alpha=0.0125$				
				Q _{MH2}	$\alpha=0.05$	Q _{MH2}	$\alpha=0.0125$			
1.2) Relación entre actividades y objetivos	-0.68	88.87	0.00	44.73	0.00	26.70	0.000	-0.53	Severo	
2.1) Características de los alumnos consideradas en la unidad	0.25	68.31	0.00	20.35	0.00	9.60	0.002	0.34	Severo*	
3.1) Calidad de la evaluación	-0.37	13.91	0.00	12.64	0.01	10.11	0.002	-0.24	Moderado	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.40	22.42	0.00	23.25	0.00	18.90	0.000	-0.27	Severo	
6.1) Calidad del inicio de la clase	-0.43	52.23	0.00	23.73	0.00	10.94	0.001	-0.37	Severo	
6.2) Calidad del cierre de la clase	-0.39	31.73	0.00	25.34	0.00	16.11	0.000	-0.30	Severo	
7.1) Calidad de las explicaciones	0.24	80.01	0.00	16.58	0.00	12.24	0.001	0.33	Severo*	
7.4) Indicador por asignatura o área	0.88	428.35	0.00	102.95	0.00	60.15	0.000	0.95	Severo*	

*DIF contra el grupo de referencia. GR=Educación Especial (EE) vs GF = Educación Media (EM)

El último análisis pareado de esta variable, por medio del estadístico SMD/σ_i , enfrentó a los grupos Educación Especial y Educación de Adultos, arrojando dos indicadores con DIF moderado (2.2 y 6.2) y cinco indicadores con DIF severo (1.2, 5.1, 5.3, 6.1 y 6.3) contra el grupo focal; y tres indicadores con DIF moderado (1.1, 5.2 y 7.1) y dos con DIF severo (2.1 y 7.4) contra el grupo de referencia.

El resumen para la comparación de los grupos Educación Especial y Educación de Adultos muestra tres indicadores con DIF considerando la confluencia de resultados en los análisis GMH (pareado y multigrupo) y SMD.

Ver Tabla12.

Tabla 12
Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación de Adultos

Indicadores	IMPACTO		DIF						Clasificación del DIF		
	D de Cohen	d	GMH Pareado		GMH Multigrupo					SMD	
			Q _{MH2}	$\alpha=0.05$	Global $\alpha=0.05$		Pareado $\alpha=0.0125$				SMD/ σ_i
					Q _{MH2}	$\alpha=0.05$	Q _{MH2}	$\alpha=0.0125$			
1.2) Relación entre actividades y objetivos	-0.69	21.21	0.00	44.73	0.00	2.95	0.086	-0.42	Severo		
2.1) Características de los alumnos consideradas en la unidad	0.27	49.92	0.00	20.35	0.00	15.95	0.000	0.42	Severo*		
7.4) Indicador por asignatura o área	0.50	60.90	0.00	102.95	0.00	18.99	0.000	0.61	Severo*		

*DIF contra el grupo de referencia. GR=Educación Especial (EE) vs GF = Educación de Adultos (EA)

4.2.5 Asignatura

4.2.5.1 Segundo Ciclo Básico

La evaluación de impacto, por medio del estadístico d de Cohen mostró, por medio de comparaciones pareadas entre cada uno de los grupos focales y el grupo de referencia, en este caso Matemática.

Evaluación de Impacto

La evaluación de impacto, por medio del cálculo de la d de Cohen, mostró que el ítem 1.1 presenta una magnitud media contra los grupos focales Educación Tecnológica ($d= -0.6$), Inglés ($d= -0.72$) y Religión Evangélica ($d= -0.71$); el ítem 1.3 presentó una magnitud media contra el grupo Historia Geografía y Ciencias Sociales ($d= -0.71$) y una

magnitud grande contra el grupo Religión Católica ($d= -1.11$). El indicador 2.2 mostró un impacto medio contra el grupo Educación Física ($d= -0.58$). Del mismo modo, los indicadores 3.1 y 3.2 presentaron magnitudes grandes contra los grupos Artes Visuales ($d= -1.39$ y $d= -1.24$ respectivamente) y Educación Tecnológica ($d= -0.83$ para ítem 3.1 y $d= -0.97$ para 3.2), al mismo tiempo, el indicador 3.1 presenta una magnitud moderada contra Artes Musicales ($d= -0.53$) y Religión Católica ($d= -0.52$); los indicadores 3.2 y 4.1 también presentan impacto medio contra Religión Católica, con valores $d= -0.76$ y $d= -0.51$ respectivamente.

El indicador 5.1 presenta una $d= -0.51$ contra el grupo Educación Tecnológica. Del mismo modo, el indicador 5.3 presenta un valor $d= -0.56$ contra el grupo Artes Musicales y $d= -0.50$ contra Artes Visuales. Por su partes, el indicador 5.2 presenta magnitudes de impacto moderadas en contra de los grupos Historia, Geografía y Ciencias Sociales ($d= -0.69$) y Religión Católica ($d= -0.70$), y magnitudes grandes contra los grupos Artes Visuales ($d= -1.05$), Educación Tecnológica ($d= -0.81$) Religión Evangélica ($d= -0.98$).

El indicador 6.1 presenta magnitudes grandes de impacto contra todos los grupos, con la única excepción de Lenguaje y Comunicación. El indicador 6.2 presenta una $d= -1.22$ en contra de Educación Física; y el indicador 6.3 presenta magnitudes medias contra Artes Visuales ($d= -0.68$), Religión Católica ($d= -0.74$) y Religión Evangélica ($d= -0.79$).

En cuanto a los indicadores de la séptima dimensión del portafolio, es posible apreciar un tamaño de impacto medio $d= -0.72$ del indicador 7.1 en contra de la asignatura de Historia, Geografía y Ciencias Sociales, y magnitudes grandes contra Educación Física ($d= -0.94$) e inglés ($d= -0.90$). Asimismo, el indicador 7.2 presenta tamaños medianos de impacto en contra de los grupos Inglés ($d= -0.73$), Religión

Católica ($d= -0.60$) y Religión Evangélica ($d= -0.54$). El indicador 7.4 también presenta una $d= -0.54$ en la comparación contra la asignatura Inglés.

Los indicadores 7.2 y 7.4 son los únicos que presentan magnitudes grandes del estadístico (d), con valores positivos, lo que denota la presencia de impacto a favor del grupo focal (o contra el grupo de referencia). El indicador 7.2 muestra un valor $d= 0.98$ a favor de Lenguaje y Comunicación, asimismo, el indicador 7.4 presenta un valor $d= 0.80$ en favor de la asignatura Educación Tecnológica.

Los indicadores que presentan magnitudes de impacto pequeñas y/o despreciables para todas las comparaciones realizadas sobre esta variable son: 1.2, 2.1, 2.3, 4.2, y 7.3.

Análisis DIF

El Test de Mantel, aplicado sobre la muestra de mayor tamaño ($N=10000$), arrojó presencia de DIF en 13 ítems, correspondientes a los indicadores 1.1, 1.3, 2.2, 2.3, 3.1, 3.2, 4.2, 5.2, 6.1, 6.2, 7.3 y 7.4, para los grupos Artes Musicales-Matemática, para un valor $\alpha=0.05$. Al corregir el nivel de significancia por medio de la fórmula de Bonferroni ($\alpha=0.00556$), el indicador 6.2 no presenta asociación significativa para la comparación de estos grupos.

La prueba para los grupos Artes Visuales-Matemática, muestra 16 ítems con asociación, correspondientes a los indicadores 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2, 4.2, 5.1, 5.2, 5.3, 6.1, 6.3, 7.1 y 7.4. Luego de aplicar la corrección de Bonferroni los indicadores 1.1, 1.2 y 2.2 no presentan asociación para el valor $\alpha=0.00556$.

El análisis para el par de grupos Historia, Geografía y Ciencias Sociales-Matemática, presenta 12 ítems con DIF (1.1, 1.2, 1.3, 2.2, 2.3, 3.2, 4.1, 4.2, 5.2, 5.3, 6.1

y 7.1). Los indicadores 4.2 y 5.3 no presentan DIF luego de corregir el nivel de significancia.

Los indicadores 1.1, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2, 4.2, 5.1, 6.1, 6.2, 6.3, 7.1, 7.3 y 7.4, muestran presencia de asociación para la comparación de los grupos Educación Física y Matemática. No presentaron asociación para $\alpha=0.00556$, los indicadores 2.3, 3.2, 6.3 y 7.4.

11 indicadores arrojaron como resultado la presencia de DIF entre los grupos Educación Tecnológica y Matemática (1.1, 2.1, 3.1, 3.2, 5.1, 5.2, 5.3, 6.1, 7.1, 7.2 y 7.4). Sólo el indicador 1.1 dejó de mostrar DIF luego de aplicar la corrección de Bonferroni.

12 indicadores presentaron DIF para los grupos Inglés-Matemática (1.1, 1.2, 2.1, 2.2, 3.2, 4.1, 4.2, 6.1, 7.1, 7.2, 7.3 y 7.4), y no hubo variación luego de ajustar el nivel de significancia.

En la comparación de los grupos Lenguaje y Comunicación-Matemática se observan 16 indicadores que presentan DIF (1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1, 5.2, 5.3, 6.1, 6.2, 6.3, 7.1, 7.2, 7.3 y 7.4). Para $\alpha=0.00556$ los indicadores 3.1, 5.3 y 7.4 no mostraron asociación.

15 ítems, correspondientes a los indicadores 1.1, 1.2, 1.3, 2.1, 2.2, 3.1, 3.2, 4.2, 5.2, 5.3, 6.1, 6.3, 7.1, 7.2 y 7.4, presentaron DIF para el análisis entre los grupos Religión Católica-Matemática, de los cuales, los indicadores 2.1, 4.2 y 7.4 no mostraron asociación al aplicar la corrección de Bonferroni. Al mismo tiempo, 11 indicadores muestran DIF entre los grupos Religión Evangélica-Matemática (1.1, 1.2, 1.3, 2.1, 2.2, 5.2, 6.1, 6.2, 6.3, 7.2 y 7.4), variando únicamente el indicador 7.2 luego de ajustar el valor de α .

El procedimiento basado en el método de estandarización arrojó los siguientes resultados. Al comparar las asignaturas de Artes Musicales (grupo focal) vs Matemática, el estadístico SMD/σ_i mostró 5 indicadores con asociación severa (3.1, 3.2, 5.2, 5.3 y 6.1) y 2 indicadores con asociación moderada (6.2 y 7.3) en contra del grupo focal. También se puede observar la presencia de asociación severa en 4 indicadores (1.3, 2.3, 4.2 y 7.4) y asociación moderada en otros 2 indicadores (1.1 y 2.2) en favor del grupo de referencia.

Tabla 13
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Artes Musicales

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	d	GMH		SMD	SMD/ σ_i	
			$\alpha=0.05$	$\alpha=0.0055$			
		Q_{MH2}	p	Q_{MH2}	p		
1.1) Formulación y coherencia entre los objetivos	0.08	12.91	0.00	10.38	0.001	0.25	Moderado*
1.3) Calidad de la secuencia entre las	0.28	41.57	0.00	35.66	0.000	0.47	Severo*
2.2) Análisis de la unidad realizada	0.09	13.53	0.00	11.82	0.001	0.24	Moderado*
2.3) Análisis de una clase	0.21	23.72	0.00	22.38	0.000	0.35	Severo*
3.1) Calidad de la evaluación	-0.53	18.35	0.00	20.58	0.000	-0.37	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.49	10.98	0.00	12.79	0.000	-0.30	Severo
4.2) Calidad de la Retroalimentación	0.11	15.95	0.00	13.88	0.000	0.27	Severo*
5.2) Promoción de la participación de todos los alumnos	-0.43	11.91	0.00	14.35	0.000	-0.32	Severo
5.3) Supervisión de las actividades	-0.56	26.69	0.00	26.63	0.000	-0.51	Severo
6.1) Calidad del inicio de la clase	-1.77	207.05	0.00	216.75	0.000	-1.64	Severo
7.3) Calidad de la retroalimentación a los alumnos	-0.38	6.53	0.01	8.24	0.000	-0.23	Moderado
7.4) Indicador por asignatura o área	0.44	63.66	0.00	65.45	0.000	0.61	Severo*

Nota: Matemática = Grupo de referencia; Artes Musicales = Grupo Focal

*DIF contra el grupo de referencia

Entre los grupos Artes Visuales (GF) y Matemática, al aplicar el método de estandarización se aprecia una asociación severa en 7 indicadores (3.1, 3.2, 5.1, 5.2, 5.3, 6.1 y 6.3) en contra del grupo focal. Además se pueden observar 4 indicadores (1.3, 2.1,

2.3 y 7.4) cuya magnitud de asociación es severa y otros 4 indicadores (1.2, 2.2, 4.2 y 7.1) con asociación moderada en favor del grupo focal.

Tabla 14
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Artes Visuales

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha=0.05$	$\alpha=0.0055$	$\alpha=0.05$	$\alpha=0.0055$		
<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>			
1.3) Calidad de la secuencia entre las	0.02	23.35	0.00	18.53	0.000	0.47	Severo*
2.1) Características de los alumnos consideradas en la unidad	0.04	20.07	0.00	14.33	0.000	0.02	Severo*
2.3) Análisis de una clase	-0.01	16.16	0.00	12.54	0.000	0.35	Severo*
3.1) Calidad de la evaluación	-1.39	131.77	0.00	140.67	0.000	-0.37	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.24	88.04	0.00	103.66	0.000	-0.30	Severo*
4.2) Calidad de la Retroalimentación	-0.14	12.12	0.00	7.79	0.005	0.27	Moderado*
5.1) Promoción de la focalización de los alumnos en las actividades	-0.49	7.11	0.01	7.95	0.005	0.00	Severo
5.2) Promoción de la participación de todos los alumnos	-1.05	65.92	0.00	67.63	0.000	-0.32	Severo
5.3) Supervisión de las actividades	-0.50	9.97	0.00	9.71	0.002	-0.51	Severo
6.1) Calidad del inicio de la clase	-1.05	47.25	0.00	43.51	0.000	-1.64	Severo
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.68	8.61	0.00	8.71	0.003	-0.01	Severo
7.1) Calidad de las explicaciones	-0.16	7.73	0.01	10.72	0.001	-0.14	Moderado
7.4) Indicador por asignatura o área	-0.06	15.84	0.00	18.68	0.000	0.61	Severo*

Nota: Matemática = Grupo de referencia; Artes Visuales = Grupo Focal

*DIF contra el grupo de referencia

El análisis que compara los grupos Historia, Geografía y Cs. Sociales (GF) versus Matemática, arrojó valores del estadístico SMD/ σ_i que indican la presencia de DIF severo en 4 indicadores (1.3, 5.2, 6.1 y 7.1), DIF moderado en el indicador 5.3, contra el grupo focal. Al mismo tiempo, 4 ítems presentaron asociación moderada (1.1, 1.2, 2.2 y 4.1) y 1 indicador (2.3) con asociación severa a favor del grupo focal.

SMD/ σ_i para los grupos Educación Física (GF) y Matemática, mostró 4 indicadores con DIF severo (2.2, 6.1, 6.2 y 7.1) y 1 con DIF moderado (2.1) en contra

del grupo focal. También se pueden apreciar 4 indicadores con DIF severo (1.1, 1.3, 3.1, 4.2 y 7.3) y 2 indicadores con DIF moderado (3.2 y 7.4) a favor del grupo focal.

Tabla 15

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Historia, Geografía y Ciencias Sociales

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
	<i>d</i>	$\alpha=0.05$		$\alpha=0.0055$		SMD/σ_i	
		Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>		
1.1) Formulación y coherencia entre los objetivos	-0.08	12.95	0.00	13.73	0.000	0.22	Moderado*
1.2) Relación entre actividades y objetivos	-0.17	8.03	0.00	8.21	0.004	0.19	Moderado*
1.3) Calidad de la secuencia entre las	-0.71	48.53	0.00	43.51	0.000	-0.35	Severo
2.2) Análisis de la unidad realizada	-0.13	9.27	0.00	10.41	0.001	0.17	Moderado*
2.3) Análisis de una clase	-0.03	22.45	0.00	23.70	0.000	0.28	Severo*
4.1) Atribución de los resultados de aprendizaje	-0.09	14.84	0.00	17.35	0.000	0.22	Moderado*
5.2) Promoción de la participación de todos los alumnos	-0.69	63.53	0.00	57.26	0.000	-0.45	Severo
6.1) Calidad del inicio de la clase	-1.08	185.52	0.00	173.02	0.000	-0.87	Severo
7.1) Calidad de las explicaciones	-0.72	45.51	0.00	38.69	0.000	-0.41	Severo

Nota: Matemática = Grupo de referencia; Historia, Geografía y Ciencias Sociales = Grupo Focal

*DIF contra el grupo de referencia

Tabla 16

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Educación Física

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
	<i>d</i>	$\alpha=0.05$		$\alpha=0.0055$		SMD/σ_i	
		Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>		
1.1) Formulación y coherencia entre los objetivos	-0.06	21.42	0.00	21.20	0.000	0.27	Moderado*
1.3) Calidad de la secuencia entre las	0.14	72.55	0.00	77.27	0.000	0.52	Severo*
2.1) Características de los alumnos consideradas en la unidad	-0.44	10.04	0.00	9.23	0.002	-0.17	Moderado
2.2) Análisis de la unidad realizada	-0.58	28.53	0.00	27.27	0.000	-0.27	Severo*
3.1) Calidad de la evaluación	-0.02	18.90	0.00	18.87	0.000	0.26	Severo*
4.2) Calidad de la Retroalimentación	-0.05	20.41	0.00	26.58	0.000	0.28	Severo*
6.1) Calidad del inicio de la clase	-1.50	333.19	0.00	349.70	0.000	-1.30	Severo
6.2) Calidad del cierre de la clase	-1.22	217.21	0.00	224.07	0.000	-0.91	Severo
7.1) Calidad de las explicaciones	-0.94	116.45	0.00	122.25	0.000	-0.68	Severo
7.3) Calidad de la retroalimentación a los alumnos	0.28	103.75	0.00	108.83	0.000	0.58	Severo*

Nota: Matemática = Grupo de referencia; Educación Física = Grupo Focal

*DIF contra el grupo de referencia

El estadístico SMD/σ_i para los grupos Educación Tecnológica (GF) y Matemática, muestra 7 indicadores con DIF severo (1.1, 3.1, 3.2, 5.1, 5.2, 5.3 y 6.1) en contra del grupo focal. Por contraparte, se pueden apreciar 3 indicadores con presencia de DIF severo (7.1, 7.2 y 7.4) y 1 indicador con DIF moderado (2.1) en contra del grupo de referencia. SMD/σ_i para Inglés (GF) y Matemática, mostró 4 indicadores con DIF severo (1.1, 6.1, 7.1 y 7.2) y 2 indicadores con DIF moderado (7.3 y 7.4) en contra del grupo focal. El indicador 2.2 presenta DIF severo, y los indicadores 2.1 y 4.1, DIF moderado en contra del grupo de referencia.

Tabla 17
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Educación Tecnológica

Indicadores	IMPACTO		DIF				Clasificación del DIF	
	D de Cohen	d	GMH		GMH			SMD
			$\alpha=0.05$	p	$\alpha=0.0055$	p		
		Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i		
1.1) Formulación y coherencia entre los objetivos	-0.60	7.96	0.00	6.60	0.010	-0.27	Severo	
2.1) Características de los alumnos consideradas en la unidad	-0.04	5.61	0.02	7.79	0.005	0.21	Moderado*	
3.1) Calidad de la evaluación	-0.83	26.31	0.00	23.04	0.000	-0.52	Severo	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.97	38.62	0.00	34.68	0.000	-0.62	Severo	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.51	9.72	0.00	8.95	0.003	-0.37	Severo	
5.2) Promoción de la participación de todos los alumnos	-0.81	32.72	0.00	28.23	0.000	-0.59	Severo	
5.3) Supervisión de las actividades	-1.09	49.05	0.00	32.97	0.000	-1.01	Severo	
6.1) Calidad del inicio de la clase	-1.26	54.66	0.00	50.29	0.000	-1.01	Severo	
7.1) Calidad de las explicaciones	0.22	23.64	0.00	23.83	0.000	0.53	Severo*	
7.2) Calidad de las preguntas	0.38	40.65	0.00	42.78	0.000	0.68	Severo*	
7.4) Indicador por asignatura o área	0.80	92.19	0.00	95.40	0.000	1.13	Severo*	

Nota: Matemática = Grupo de referencia; Educación Tecnológica = Grupo Focal

*DIF contra el grupo de referencia

Al realizar la comparación entre los grupos Lenguaje y Comunicación (GF) y Matemática, aplicando el método de estandarización, se aprecia una asociación severa en 3 indicadores (3.2, 5.2 y 6.2) y asociación moderada en 2 indicadores (6.1 y 7.2) en

contra del grupo focal. Además, 2 indicadores mostraron una magnitud de asociación moderada (4.2 y 7.4) en contra del grupo Matemática (Tabla 19).

Tabla 18
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Inglés

Indicadores	IMPACTO	DIF				Clasificación del DIF	
	D de Cohen	GMH		GMH			SMD
		$\alpha=0.05$	$\alpha=0.0055$	$\alpha=0.0055$	$\alpha=0.0055$		
<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>	SMD/ σ_i		
1.1) Formulación y coherencia entre los objetivos	-0.72	38.47	0.00	44.26	0.000	-0.42	Severo
2.1) Características de los alumnos consideradas en la unidad	0.05	34.58	0.00	37.69	0.000	0.27	Severo*
2.2) Análisis de la unidad realizada	0.33	103.03	0.00	108.88	0.000	0.59	Severo*
4.1) Atribución de los resultados de aprendizaje	0.09	46.44	0.00	49.04	0.000	0.40	Severo*
6.1) Calidad del inicio de la clase	-1.14	184.65	0.00	182.77	0.000	-0.92	Severo
7.1) Calidad de las explicaciones	-1.00	128.78	0.00	133.52	0.000	-0.73	Severo
7.2) Calidad de las preguntas	-0.73	68.20	0.00	69.85	0.000	-0.46	Severo
7.3) Calidad de la retroalimentación a los alumnos	-0.46	7.17	0.01	9.22	0.002	-0.17	Moderado
7.4) Indicador por asignatura o área	-0.54	14.13	0.00	14.39	0.000	-0.24	Moderado

Nota: Matemática = Grupo de referencia; Inglés = Grupo Focal

*DIF contra el grupo de referencia

Tabla 19
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Lenguaje y Comunicación

Indicadores	IMPACTO	DIF				Clasificación del DIF	
	D de Cohen	GMH		GMH			SMD
		$\alpha=0.05$	$\alpha=0.0055$	$\alpha=0.0055$	$\alpha=0.0055$		
<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>	SMD/ σ_i		
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.35	36.12	0.00	31.87	0.000	-0.27	Severo*
4.2) Calidad de la Retroalimentación	0.09	9.42	0.00	16.97	0.000	0.17	Moderado
5.2) Promoción de la participación de todos los alumnos	-0.45	77.39	0.00	73.24	0.000	-0.40	Severo*
6.1) Calidad del inicio de la clase	-0.27	24.22	0.00	23.20	0.000	-0.22	Moderado
6.2) Calidad del cierre de la clase	-0.36	42.74	0.00	40.41	0.000	-0.28	Severo*
7.1) Calidad de las explicaciones	-0.31	33.16	0.00	30.44	0.000	-0.24	Moderado
7.2) Calidad de las preguntas	0.98	404.84	0.00	426.17	0.000	1.06	Severo*
7.4) Indicador por asignatura o área	0.07	5.29	0.02	7.52	0.006	0.17	Moderado*

Nota: Matemática = Grupo de referencia; Lenguaje y Comunicación = Grupo Focal

*DIF contra el grupo de referencia

El análisis entre los grupos Religión Católica (GF) y Matemática mostró 6 indicadores con presencia de DIF severo (1.3, 3.2, 5.2, 5.3, 6.1 y 6.3) y uno con DIF moderado (7.2) en contra del grupo focal. Igualmente, 4 indicadores presentan DIF severo (1.1, 1.2, 2.2 y 7.1) y un indicador presenta DIF moderado (7.4) a favor del grupo focal.

Tabla 20
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Religión Católica

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha=0.05$	$\alpha=0.0055$	$\alpha=0.0055$	$\alpha=0.0055$		
d	Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i		
1.1) Formulación y coherencia entre los objetivos	-0.08	18.93	0.00	12.37	0.000	0.34	Severo*
1.2) Relación entre actividades y objetivos	-0.14	21.02	0.00	16.97	0.000	0.32	Severo*
1.3) Calidad de la secuencia entre las	-1.11	82.93	0.00	103.49	0.000	-0.64	Severo
2.2) Análisis de la unidad realizada	-0.11	19.63	0.00	20.87	0.000	0.26	Severo*
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.76	17.38	0.00	24.74	0.000	-0.29	Severo*
5.2) Promoción de la participación de todos los alumnos	-0.70	34.75	0.00	44.97	0.000	-0.41	Severo
5.3) Supervisión de las actividades	-0.46	10.96	0.00	13.24	0.000	-0.34	Severo
6.1) Calidad del inicio de la clase	-1.06	75.43	0.00	84.74	0.000	-0.74	Severo
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.74	12.68	0.00	17.84	0.000	-0.41	Severo
7.1) Calidad de las explicaciones	0.05	50.14	0.00	40.71	0.000	0.42	Severo*
7.2) Calidad de las preguntas	-0.60	15.90	0.00	26.06	0.000	-0.19	Moderado
7.4) Indicador por asignatura o área	-0.22	10.60	0.00	5.23	0.022	0.21	Moderado*

Nota: Matemática = Grupo de referencia; Religión Católica = Grupo Focal

*DIF contra el grupo de referencia

La comparación entre Religión Evangélica (GF) y Matemática muestra 4 indicadores con DIF severo (1.1, 5.2, 6.1 y 6.3) y uno con DIF moderado (5.3) en contra del grupo focal. Conjuntamente, se observan 5 indicadores con presencia de DIF severo (1.3, 2.1, 2.2, 6.2 y 7.4) y uno con DIF moderado (1.2) en contra del grupo de referencia.

Tabla 21
Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Religión Evangélica

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha=0.05$	$\alpha=0.0055$	$\alpha=0.0055$	$\alpha=0.0055$		
<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>	SMD/σ_i		
1.1) Formulación y coherencia entre los objetivos	-0.71	10.12	0.00	8.08	0.005	-0.43	Severo
1.2) Relación entre actividades y objetivos	-0.12	12.46	0.00	12.04	0.001	0.20	Moderado
1.3) Calidad de la secuencia entre las	-0.07	11.36	0.00	11.70	0.001	0.26	Severo*
2.1) Características de los alumnos consideradas en la unidad	0.10	19.34	0.00	20.64	0.000	0.33	Severo*
2.2) Análisis de la unidad realizada	0.21	30.63	0.00	36.50	0.000	0.48	Severo*
5.2) Promoción de la participación de todos los alumnos	-0.98	55.37	0.00	52.52	0.000	-0.77	Severo
6.1) Calidad del inicio de la clase	-0.96	23.10	0.00	18.87	0.000	-0.74	Severo
6.2) Calidad del cierre de la clase	0.15	27.69	0.00	32.83	0.000	0.46	Severo*
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.79	12.78	0.00	10.63	0.001	-0.54	Severo
7.4) Indicador por asignatura o área	0.13	29.56	0.00	36.52	0.000	0.46	Severo*

Nota: Matemática = Grupo de referencia; Religión Evangélica = Grupo Focal

*DIF contra el grupo de referencia

4.2.5.2 Educación Media

Evaluación de Impacto

Al implementar el análisis de impacto para la variable asignatura, en el grupo Educación Media, se identificaron magnitudes grandes del estadístico d de Cohen en diez indicadores. En la primera dimensión del portafolio el indicador 1.1 presentó una magnitud de .92 contra la asignatura Inglés, y el indicador 1.3 presentó una magnitud de .81 contra la asignatura Religión Católica.

No se detectaron indicadores con magnitudes grandes en la segunda dimensión del portafolio, siendo el valor más alto contra la asignatura de Educación Física, con una $d = -.55$.

Los indicadores 3.1 y 3.2 mostraron de forma consistente Impacto de magnitudes grandes contra las asignaturas Artes Visuales (d igual a -1.73 y -1.8 respectivamente), Educación tecnológica (d igual a -1.2 y -1.45 respectivamente) y Religión Católica (d igual a $-.85$ y -1.11 respectivamente).

El análisis para los indicadores de la cuarta y quinta dimensión del portafolio muestra que los indicadores 4.1, 4.2, 5.1 y 5.3 no presentaron impacto de magnitudes grandes. El indicador 5.2 mostró una $d = -1.0$ contra la asignatura Artes Visuales.

El indicador 6.1 mostró impacto de magnitudes grandes contra las asignaturas Artes Musicales y Educación Física ($d = -1.05$ y $d = -0.86$ respectivamente), mientras el indicador 6.2 presentó una $d = -1.06$ contra Educación Física. El indicador 6.3 muestra Impacto contra Religión Católica ($d = -0.96$).

En la séptima dimensión del portafolio, el indicador 7.4 presentó impacto de magnitud en contra del grupo Química, con una $d = -1.02$. Por otra parte, los indicadores

7.2 y 7.4 mostraron Impacto contra el grupo de referencia frente a los grupos Lenguaje ($d=1.02$) y Educación Tecnológica ($d=1.46$) respectivamente.

Análisis DIF

Los análisis para la comparación pareada de los grupos Matemática (grupo de referencia) y el grupo focal Artes Musicales mostraron los siguientes resultados. El procedimiento de Mantel Haenszel para el nivel de significación $\alpha= 0.05$ mostró 13 ítems con DIF, excluyéndose los indicadores 1.2, 2.1, 4.1, 5.1, 6.3, 7.1 y 7.2. Al aplicar el procedimiento sobre los indicadores detectados con DIF en el primer análisis, con el nivel de significación ajustado a $\alpha= 0.0045$ de acuerdo a la corrección de Bonferroni, presentaron DIF los indicadores 1.3, 2.2, 3.1, 3.2, 6.1 y 7.4.

El procedimiento SMD mostró DIF severo contra el grupo focal en los indicadores 3.1, 3.2, 5.2, 5.3 y 6.1, DIF moderado en los indicadores 6.2, 6.3 y 7.3. También se observó DIF contra el grupo de referencia de magnitud moderada en los indicadores 4.2 y 7.2, y de magnitud severa en los indicadores 1.3, 2.2 y 7.4. La convergencia de los análisis se presenta en la Tabla 22.

Tabla 22
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Artes Musicales

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha= 0.05$	$\alpha= 0.0045$	$\alpha= 0.0045$	$\alpha= 0.0045$		
d	Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i		
1.3) Calidad de la secuencia entre las	0.28	41.57	0.00	31.14	0.000	0.53	Severo*
2.2) Análisis de la unidad realizada	0.17	13.53	0.00	21.98	0.000	0.38	Severo*
3.1) Calidad de la evaluación	-0.54	18.35	0.00	15.41	0.000	-0.39	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.63	10.98	0.00	21.33	0.000	-0.39	Severo
6.1) Calidad del inicio de la clase	-1.05	207.05	0.00	69.48	0.000	-0.89	Severo
7.4) Indicador por asignatura o área	0.37	63.66	0.00	46.92	0.000	0.58	Severo*

Nota: Matemática = Grupo de referencia; Artes Musicales = Grupo Focal

*DIF contra el grupo de referencia

La comparación de los grupos Matemática vs Artes Visuales mostró, para el primer análisis GMH, que solamente 4 indicadores estaban libres de DIF (1.1, 4.1, 5.1 y 6.2), los 16 restantes fueron sometidos al segundo análisis, con el nivel alpha corregido, y el resultado para estas condiciones de análisis mostró que los indicadores 1.3, 2.1, 3.1, 3.2, 4.2, 5.2, 5.3, 6.1, 6.3, 7.1, 7.2, 7.3 y 7.4 presentaban DIF.

El procedimiento SMD mostró DIF moderado contra el grupo focal en los indicadores 1.1 y 5.1, y DIF severo en los indicadores 3.1, 3.2, 5.2, 5.3, 6.1 y 6.3. Al mismo tiempo, se observa la presencia de DIF contra el grupo de referencia de magnitud moderada en los indicadores 2.2 y 2.3, y de magnitud severa en los indicadores 1.3, 2.1, 4.2, 7.1, 7.2, 7.3 y 7.4.

De acuerdo al criterio de convergencia de los análisis, 13 indicadores presentan DIF severo para la comparación de los grupos Matemática y Artes Visuales, según se muestra en la Tabla 23.

La aplicación del procedimiento generalizado de Mantel-Haenszel, para los grupos Matemática y Biología arrojó que los indicadores 1.2, 1.3, 3.2, 4.1, 5.2, 5.3, 7.1, 7.2 y 7.4 presentan DIF al nivel de significación 0.05. Al volver a analizar estos indicadores con el nivel $\alpha = 0.0045$, los indicadores 5.3 y 7.4 dejaron de mostrar asociación.

El método de estandarización mostró DIF moderado contra el grupo focal, en los indicadores 2.1, 6.1 y 7.4, y DIF severo en los indicadores 1.2, 3.2, 5.1, 5.2 y 5.3. Los indicadores 1.3, 4.1, 7.1 y 7.2 presentaron DIF severo contra el grupo de referencia. Ver Tabla 24 para observar la coincidencia de los análisis.

Tabla 23*Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Artes Visuales*

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	d	GMH		SMD		
			GMH $\alpha=0.05$	GMH $\alpha=0.0045$		SMD/ σ_i	
		Q _{MH2}	p	Q _{MH2}	p		
1.3) Calidad de la secuencia entre las	0.25	50.57	0.00	48.37	0.000	0.58	Severo*
2.1) Características de los alumnos consideradas en la unidad	0.13	29.25	0.00	28.87	0.000	0.42	Severo*
3.1) Calidad de la evaluación	-1.73	157.62	0.00	167.71	0.000	-1.49	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.80	149.14	0.00	162.77	0.000	-1.43	Severo
4.2) Calidad de la Retroalimentación	-0.10	14.01	0.00	13.52	0.000	0.26	Severo*
5.2) Promoción de la participación de todos los alumnos	-1.00	55.18	0.00	61.62	0.000	-0.80	Severo
5.3) Supervisión de las actividades	-0.50	9.58	0.00	11.89	0.001	-0.39	Severo
6.1) Calidad del inicio de la clase	-0.55	8.17	0.00	12.14	0.001	-0.34	Severo
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.59	7.90	0.00	11.06	0.001	-0.35	Severo
7.1) Calidad de las explicaciones	0.34	66.05	0.00	58.00	0.000	0.62	Severo*
7.2) Calidad de las preguntas	0.17	29.01	0.00	22.26	0.000	0.41	Severo*
7.3) Calidad de la retroalimentación a los alumnos	-0.01	15.54	0.00	10.80	0.001	0.30	Severo*
7.4) Indicador por asignatura o área	0.18	34.96	0.00	26.60	0.000	0.48	Severo*

Nota: Matemática = Grupo de referencia; Artes Visuales = Grupo Focal

*DIF contra el grupo de referencia

Tabla 24*Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Biología*

Indicadores	IMPACTO		DIF				Clasificación del DIF
	D de Cohen	d	GMH		SMD		
			GMH $\alpha=0.05$	GMH $\alpha=0.0045$		SMD/ σ_i	
		Q _{MH2}	p	Q _{MH2}	p		
1.2) Relación entre actividades y objetivos	-0.73	21.17	0.00	18.33	0.000	-0.40	Severo
1.3) Calidad de la secuencia entre las	0.00	5.97	0.01	8.13	0.004	0.28	Severo*
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.78	26.07	0.00	22.13	0.000	-0.51	Severo
4.1) Atribución de los resultados de aprendizaje	0.13	24.06	0.00	31.04	0.000	0.48	Severo*
5.2) Promoción de la participación de todos los alumnos	-0.59	15.81	0.00	12.56	0.000	-0.41	Severo
7.1) Calidad de las explicaciones	0.41	59.07	0.00	65.18	0.000	0.65	Severo*
7.2) Calidad de las preguntas	0.35	45.90	0.00	71.11	0.000	0.54	Severo*

Nota: Matemática = Grupo de referencia; Biología = Grupo Focal

*DIF contra el grupo de referencia

El estadístico generalizado de Mantel-Haenszel, para los grupos Matemática y Educación Física, al nivel de significación $\alpha = 0.05$, muestra que presentan DIF los indicadores 1.2, 1.3, 2.2, 4.2, 5.1, 5.3, 6.1, 6.2, 7.1 y 7.3. Al analizar dichos indicadores al nivel de significación ajustado, siete de los indicadores evaluados presentaron DIF: 1.3, 2.2, 5.1, 6.1, 6.2, 7.1 y 7.4.

El método SMD muestra que el indicador 2.2 presenta DIF moderado y los indicadores 6.1, 6.2 y 7.1 presentan DIF severo contra el grupo focal. Los indicadores 5.3 y 7.2 muestran DIF moderado contra el grupo de referencia, así como los indicadores 1.3, 5.1, 7.3 y 7.4 muestran DIF severo, también en favor del grupo focal.

Los indicadores que presentarían DIF al aplicar el criterio de convergencia entre los análisis se muestran en la Tabla 25

Tabla 25
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Educación Física

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha = 0.05$	$\alpha = 0.0045$	$\alpha = 0.0045$	$\alpha = 0.0045$		
	D	Q _{MH2}	p	Q _{MH2}	p	SMD/ σ_i	
1.3) Calidad de la secuencia entre las	0.30	63.88	0.00	59.21	0.000	0.67	Severo*
2.2) Análisis de la unidad realizada	-0.55	4.66	0.03	9.74	0.002	-0.21	Moderado
5.1) Promoción de la focalización de los alumnos en las actividades	0.19	12.87	0.00	11.16	0.001	0.34	Severo*
6.1) Calidad del inicio de la clase	-0.86	37.69	0.00	44.88	0.000	-0.65	Severo
6.2) Calidad del cierre de la clase	-1.06	72.32	0.00	72.60	0.000	-0.81	Severo
7.1) Calidad de las explicaciones	-0.61	12.84	0.00	13.56	0.000	-0.33	Severo
7.3) Calidad de la retroalimentación a los alumnos	0.34	46.26	0.00	51.06	0.000	0.69	Severo*

Nota: Matemática = Grupo de referencia; Educación Física = Grupo Focal

*DIF contra el grupo de referencia

El estadístico generalizado MH para la comparación de las asignaturas Educación Tecnológica y Matemática arrojó presencia de DIF para los indicadores 1.1,

1.3, 2.1, 3.1, 3.2, 5.3, 6.1, 7.1, 7.2 y 7.4 al nivel de significación $\alpha= 0.05$, el segundo análisis, al nivel $\alpha= 0.0045$ mostró presencia de DIF en los mismos indicadores, exceptuando el indicador 2.1 y el indicador 7.3.

El estadístico SMD/σ_i muestra DIF de magnitud moderada, contra el grupo focal, en el indicador 6.3, y DIF severo en los indicadores 1.1, 3.1, 3.2, 5.3 y 6.1. También se observa DIF moderado contra el grupo de referencia en el indicador 2.1, y DIF severo en los indicadores 1.3, 7.1, 7.2 y 7.4.

En resumen, se consideran con DIF para esta comparación, los indicadores presentados en la Tabla 26.

Tabla 26
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Educación Tecnológica

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha= 0.05$	$\alpha= 0.0045$	$\alpha= 0.0045$	$\alpha= 0.0045$		
	<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>		
1.1) Formulación y coherencia entre los objetivos	-0.79	29.17	0.00	30.01	0.000	-0.55	Severo
1.3) Calidad de la secuencia entre las	0.07	14.68	0.00	14.18	0.000	0.32	Severo*
3.1) Calidad de la evaluación	-1.20	83.86	0.00	87.77	0.000	-1.03	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.45	117.80	0.00	122.05	0.000	-1.18	Severo
6.1) Calidad del inicio de la clase	-0.77	26.77	0.00	29.30	0.000	-0.62	Severo
7.1) Calidad de las explicaciones	0.58	73.57	0.00	66.44	0.000	0.79	Severo*
7.2) Calidad de las preguntas	0.60	85.45	0.00	81.03	0.000	0.79	Severo*
7.4) Indicador por asignatura o área	1.46	233.38	0.00	226.84	0.000	1.66	Severo*

Nota: Matemática = Grupo de referencia; Educación Tecnológica = Grupo Focal

*DIF contra el grupo de referencia

El primer análisis realizado sobre los grupos Historia, Geografía y Ciencias Sociales v/s Matemática mostró 13 indicadores con DIF (1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.2, 4.1, 5.2, 6.1, 6.3, 7.1 y 7.2). De estos indicadores, solamente los indicadores 1.1 y 3.2 no

presentaron DIF luego del análisis al nivel de significación ajustado de acuerdo con la corrección de Bonferroni.

El estadístico SMD/σ_i muestra que 6 indicadores presentan DIF contra el grupo focal: 3.1 y 3.2 (DIF moderado), 1.3, 5.2, 6.1, 6.3 y 7.1 (DIF severo). También se identificaron dos indicadores con DIF moderado contra el grupo de referencia (1.1 y 2.3) y cuatro indicadores con DIF severo contra la asignatura Matemática: 1.2, 2.2, 4.1 y 7.2. La Tabla 27 resume el resultado para la comparación de estos grupos.

Tabla 27

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Historia, Geografía y Ciencias Sociales

Indicadores	IMPACTO		DIF				Clasificación del DIF	
	D de Cohen	d	GMH		GMH			SMD
			$\alpha=0.05$	p	$\alpha=0.0045$	p		
		Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i		
1.2) Relación entre actividades y objetivos	0.05	19.31	0.00	26.31	0.000	0.31	Severo*	
1.3) Calidad de la secuencia entre las	-0.64	19.40	0.00	25.16	0.000	-0.44	Severo*	
2.2) Análisis de la unidad realizada	0.45	83.80	0.00	88.40	0.000	0.66	Severo*	
2.3) Análisis de una clase	-0.02	11.34	0.00	10.39	0.001	0.17	Moderado*	
4.1) Atribución de los resultados de aprendizaje	-0.02	15.79	0.00	12.73	0.000	0.26	Severo*	
5.2) Promoción de la participación de todos los alumnos	-0.65	24.07	0.00	34.20	0.000	-0.48	Severo	
6.1) Calidad del inicio de la clase	-0.50	12.70	0.00	14.22	0.000	-0.34	Severo	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.66	23.73	0.00	31.52	0.000	-0.50	Severo	
7.1) Calidad de las explicaciones	-0.50	5.96	0.01	13.08	0.000	-0.27	Severo	
7.2) Calidad de las preguntas	0.09	32.88	0.00	24.54	0.000	0.26	Severo*	

Nota: Matemática = Grupo de referencia; Historia, Geografía y Ciencias Sociales = Grupo Focal

*DIF contra el grupo de referencia

Los resultados para la comparación de los grupos Matemática e Inglés son los siguientes. El procedimiento generalizado de Mantel-Haenszel mostró coincidencia para la mayoría de los indicadores bajo los niveles de significación $\alpha=0.05$ y $\alpha=0.0045$, detectándose 11 indicadores en el primer caso, de los cuales solo el indicador 6.2 no presentó DIF en el segundo análisis. De este modo, los indicadores identificados con DIF para este procedimiento fueron: 1.1, 2.1, 2.2, 3.1, 3.2, 4.1, 6.1, 6.3, 7.1 y 7.4.

El procedimiento de estandarización mostró presencia de DIF severo en 11 indicadores: 1.1, 3.1, 3.2, 5.2, 6.3 y 7.3 contra el grupo focal, y 1.3, 2.1, 2.2, 4.2 y 7.2 contra el grupo de referencia. También se identificó DIF moderado contra el grupo focal en el indicador 6.1.

Siete indicadores cumplen con el requisito de convergencia de análisis utilizado para concluir la presencia de DIF. Ver Tabla 28.

Tabla 28
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Inglés

Indicadores	IMPACTO		DIF				Clasificación del DIF	
	D de Cohen	d	GMH		GMH			SMD
			$\alpha=0.05$	P	$\alpha=0.0045$	p		
1.1) Formulación y coherencia entre los objetivos	-0.92	87.29	0.00	90.45	0.000	-0.26	Severo	
2.1) Características de los alumnos consideradas en la unidad	0.15	28.73	0.00	32.00	0.000	0.32	Severo*	
2.2) Análisis de la unidad realizada	0.40	74.96	0.00	78.49	0.000	0.43	Severo*	
3.1) Calidad de la evaluación	-0.57	34.40	0.00	34.94	0.000	-0.46	Severo	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.42	10.82	0.00	11.29	0.001	-0.53	Severo	
6.1) Calidad del inicio de la clase	-0.57	36.85	0.00	33.93	0.000	-0.19	Moderado*	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.36	7.38	0.01	8.17	0.004	-0.42	Severo	

Nota: Matemática = Grupo de referencia; Inglés = Grupo Focal

*DIF contra el grupo de referencia

Se observaron 14 indicadores con DIF, para la comparación de los grupos Lenguaje y Comunicación y Matemática, de acuerdo al estadístico generalizado MH, al nivel de significación $\alpha=0.05$, el procedimiento con el nivel de significación ajustado $\alpha=0.0045$, aplicado sobre los ítems identificados con DIF en este primer análisis arrojó presencia de DIF en los indicadores 1.1, 1.3, 2.1, 2.2, 3.1, 3.2, 4.2, 5.2, 6.1, 6.3, 7.2 y 7.3.

Para la comparación contra el grupo focal Lenguaje y Comunicación, el estadístico SMD/σ_i mostró presencia de DIF moderado contra el grupo focal en los indicadores 5.1 y 6.1, y contra el grupo de referencia en el indicador 4.2. También se identificó con DIF severo, contra el grupo focal, a los indicadores 1.1, 3.1, 3.2, 5.2, 6.3 y 7.3, y contra el grupo de referencia a los indicadores 1.3, 2.1, 2.2 y 7.2.

El resumen para los resultados de la comparación entre estos grupos se muestra en la Tabla 29.

Tabla 29
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Lenguaje y Comunicación

Indicadores	IMPACTO		DIF				Clasificación del DIF	
	D de Cohen	d	GMH		GMH			SMD
			$\alpha=0.05$	P	$\alpha=0.0045$	p		
1.1) Formulación y coherencia entre los objetivos	-0.42	22.13	0.00	17.34	0.00	-0.31	Severo	
1.3) Calidad de la secuencia entre las	0.48	49.22	0.00	69.40	0.00	0.58	Severo*	
2.1) Características de los alumnos consideradas en la unidad	0.20	17.06	0.00	22.02	0.00	0.32	Severo*	
2.2) Análisis de la unidad realizada	0.34	27.85	0.00	39.45	0.00	0.46	Severo*	
3.1) Calidad de la evaluación	-0.53	26.10	0.00	33.02	0.00	-0.47	Severo	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.68	45.10	0.00	56.09	0.00	-0.59	Severo	
4.2) Calidad de la Retroalimentación	0.11	12.81	0.00	17.68	0.00	0.25	Moderado*	
5.2) Promoción de la participación de todos los alumnos	-0.55	27.53	0.00	32.62	0.00	-0.46	Severo	
6.1) Calidad del inicio de la clase	-0.29	12.75	0.00	10.39	0.00	-0.22	Moderado	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.46	26.79	0.00	26.04	0.00	-0.39	Severo	
7.2) Calidad de las preguntas	1.02	168.71	0.00	205.04	0.00	1.11	Severo*	
7.3) Calidad de la retroalimentación a los alumnos	-0.54	24.33	0.00	26.48	0.00	-0.43	Severo*	

Nota: Matemática = Grupo de referencia; Lenguaje y Comunicación = Grupo Focal

*DIF contra el grupo de referencia

Los resultados para la comparación de los grupos Matemática y Química arrojan, para el estadístico generalizado MH, al nivel de significación $\alpha=0.05$, presencia de DIF en los indicadores 1.1, 1.2, 2.2, 3.1, 4.1, 5.3, 6.1, 6.2, 6.3, 7.1, 7.2, 7.3 y 7.4. Al analizar estos indicadores al nivel de significancia ajustado por medio de la corrección de Bonferroni, presentan DIF los indicadores 1.1, 3.1, 5.3, 6.1, 6.3, 7.1, 7.2, 7.3 y 7.4.

La comparación de los grupos Química y Matemática con el procedimiento SMD muestra presencia de DIF severo contra el grupo focal en los indicadores 1.1, 3.1, 6.3, 6.1, 6.3, 7.3 y 7.4. También se observa DIF contra el grupo de referencia de magnitud moderada en los indicadores 1.2 y 4.1, y de magnitud severa en los indicadores 2.2, 6.2, 7.1 y 7.2.

Como se observa en la Tabla 30, nueve indicadores muestran DIF para la convergencia de los resultados obtenidos en los diferentes análisis implementados para la comparación de estos grupos.

Tabla 30
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Química

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha=0.05$	$\alpha=0.0045$	$\alpha=0.05$	$\alpha=0.0045$		
	<i>d</i>	Q_{MH2}	<i>p</i>	Q_{MH2}	<i>p</i>		
1.1) Formulación y coherencia entre los objetivos	-0.43	8.55	0.00	8.30	0.004	-0.28	Severo
3.1) Calidad de la evaluación	-0.44	7.81	0.01	9.75	0.002	-0.34	Severo
5.3) Supervisión de las actividades	-0.47	7.24	0.01	8.20	0.004	-0.39	Severo
6.1) Calidad del inicio de la clase	-0.60	19.04	0.00	25.55	0.000	-0.49	Severo
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.53	13.46	0.00	11.31	0.001	-0.41	Severo
7.1) Calidad de las explicaciones	0.56	35.73	0.00	37.11	0.000	0.70	Severo*
7.2) Calidad de las preguntas	0.50	29.65	0.00	25.93	0.000	0.62	Severo*
7.3) Calidad de la retroalimentación a los alumnos	-0.52	13.67	0.00	15.84	0.000	-0.38	Severo
7.4) Indicador por asignatura o área	-1.02	64.08	0.00	68.90	0.000	-0.89	Severo

Nota: Matemática = Grupo de referencia; Química = Grupo Focal

*DIF contra el grupo de referencia

Finalmente, la comparación de los grupos Matemática y Religión Católica muestra 10 indicadores con DIF según el resultado del primer análisis generalizado de Mantel-Haenszel ($\alpha=0.05$), nueve de los cuales confirman la presencia de DIF luego de aplicar el análisis con el nivel de significación corregido ($\alpha=0.0045$): 1.2, 1.3, 2.1, 2.2, 3.1, 3.2, 6.3, 7.1 y 7.2.

El método de estandarización muestra presencia de DIF moderado en el indicador 1.1, contra la asignatura de Religión Católica, y DIF severo contra el mismo grupo en los indicadores 1.3, 3.1, 3.2, 5.2, 5.3, 6.1 y 6.3. Al mismo tiempo, se observa presencia de DIF moderado contra el grupo de referencia en el indicador 2.1, y DIF severo en los indicadores 1.2, 2.2, 7.1 y 7.2.

Los indicadores que cumplen el criterio de convergencia en los resultados de los distintos análisis y, por tanto, se consideran con DIF, se muestran en la Tabla 31.

Tabla 31
Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Religión Católica

Indicadores	IMPACTO	DIF					Clasificación del DIF
	D de Cohen	GMH		GMH		SMD	
		$\alpha=0.05$	$\alpha=0.0045$	$\alpha=0.0045$	$\alpha=0.0045$		
d	Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i		
1.2) Relación entre actividades y objetivos	0.12	33.27	0.00	35.37	0.00	0.60	Severo*
1.3) Calidad de la secuencia entre las	-0.81	13.67	0.00	12.75	0.00	-0.40	Severo
2.1) Características de los alumnos consideradas en la unidad	-0.11	10.90	0.00	10.68	0.00	0.24	Moderado*
2.2) Análisis de la unidad realizada	0.11	28.76	0.00	27.38	0.00	0.48	Severo*
3.1) Calidad de la evaluación	-0.85	21.85	0.00	20.55	0.00	-0.57	Severo
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.11	37.14	0.00	40.28	0.00	-0.68	Severo
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.96	15.41	0.00	13.46	0.00	-0.67	Severo
7.1) Calidad de las explicaciones	0.41	58.66	0.00	66.07	0.00	0.76	Severo*
7.2) Calidad de las preguntas	0.00	14.42	0.00	20.35	0.00	0.33	Severo*

Nota: Matemática = Grupo de referencia; Religión Católica = Grupo Focal

*DIF contra el grupo de referencia

4.3 Resumen por indicador

Finalmente, al observar los resultados en función de cada indicador luego de los análisis realizados en 31 comparaciones entre los grupos focales y de referencia definidos para cada variable, y considerando el criterio de convergencia de los análisis practicados, se observa que todos los indicadores presentan DIF para alguna de las situaciones en estudio, concentrándose los casos, en primera medida, en los análisis de la variable Asignatura, y en segundo lugar, en la variable Nivel de Enseñanza.

Por otra parte, las variables Sexo y Edad presentan un único caso con evidencia de DIF, el indicador 6.1: *Calidad del inicio de la clase*. La variable Zona Geográfica no presentó indicadores con DIF. La figura n°9 resume los resultados por indicador para cada una de las comparaciones contempladas en el presente estudio.

RESUMEN ANÁLISIS DIF POR INDICADORES DEL PORTAFOLIO																																			
INDICADOR	Variables																			N° de comparaciones con DIF															
	SEXO		ZONA GEOGRÁFICA		EDAD				NIVEL DE ENSEÑANZA						ASIGNATURA																				
	H-M	R-U	E3 VS E4	E3 VS E6	E3 VS E7	E3 VS E8	PC-EP	PC-SC	PC-BM	PC-EA	EE-EP	EE-SC	EE-BM	EE-EA	SEGUNDO CICLO				EDUCACIÓN MEDIA																
														M-AM	M-AV	M-CS	M-EF	M-ET	M-I	M-L	M-RC	M-RE	M-AM	M-AV	M-B	M-EF	M-ET	M-H	M-I	M-L	M-Q	M-RC			
1.1																																		8	
1.2																																			9
1.3																																			13
2.1																																			8
2.2																																			9
2.3																																			3
3.1																																			13
3.2																																			14
4.1																																			3
4.2																																			4
5.1																																			3
5.2																																			12
5.3																																			8
6.1																																			18
6.2																																			7
6.3																																			9
7.1																																			14
7.2																																			12
7.3																																			7
7.4																																			12

Figura 9. Resumen de indicadores que presentan DIF severo.

V DISCUSIÓN Y CONCLUSIONES

5.1 Discusión

En primer lugar, es importante señalar que los análisis inferenciales arrojaron la existencia de diferencias estadísticamente significativas en los resultados obtenidos por los diferentes grupos incluidos en el presente estudio, lo que fundamenta una indagación más profunda en el origen de dichas diferencias, con el fin de determinar si éstas se deben a ventajas reales entre las habilidades de los diferentes grupos estudiados, o bien, se deben a distorsiones surgidas por características del instrumento de medición.

Dicho lo anterior, es necesario revisar los resultados atendiendo, por una parte al tipo de análisis implementado, y por otro, al tipo de decisiones que permite orientar cada caso. En primer lugar, cabe señalar que los análisis realizados con muestras grandes, para el estadístico generalizado MH, arrojaron la mayor cantidad de indicadores con DIF, al punto que, para los análisis multigrupo, fueron señalados con presencia de DIF la totalidad de los indicadores. Lo anterior se puede explicar porque, al realizar este tipo de análisis, como en las variables *Edad*, *Nivel de Enseñanza* y *Asignatura*, la prueba estadística realiza simultáneamente múltiples comparaciones, pudiendo identificar grados de asociación significativa para una o varias de las comparaciones, en cada uno de los indicadores. Sin descartar la posibilidad real de que los niveles de asociación pudieran efectivamente dar cuenta de la existencia de DIF, no parece razonablemente responsable orientar alguna decisión a partir de este único criterio.

Por otro lado, el estadístico generalizado MH no entrega una magnitud de la asociación, ni tampoco informa acerca de la dirección de la misma, lo que justifica que la interpretación se lleve a cabo a partir de la confluencia de análisis, donde el estadístico SMD/σ_i completa la información que entrega el valor de la significación estadística con una medida de la magnitud de dicha asociación y una dirección que, en

este caso, es mucho más informativa respecto al comportamiento de los indicadores que se someten a estudio.

Al detener la atención en el análisis SMD, es posible observar la presencia simultánea de indicadores que favorecen al grupo de referencia, al tiempo que otros favorecen al grupo focal. Considerando que la presencia de valores absolutos grandes en el estadístico SMD/σ_i sugiere una mayor ventaja en la respuesta al ítem, la presencia de valores negativos y positivos simultáneamente puede ser evidencia de DIF compensatorio. Lo anterior es muy importante porque, aunque puede significar que hay indicadores que favorecen a determinados grupos, también se puede relacionar con la dimensionalidad del portafolio.

Siguiendo el criterio del ETS, se centrará la atención en los indicadores que presentan magnitudes de DIF severas (ver *Figura 9*). También es posible establecer una cierta diferencia en la naturaleza de las variables en estudio, donde se observan aquellas referidas a características demográficas, tales como *Sexo*, *Zona Geográfica* y *Edad*, y otras variables referidas a elementos al desarrollo y quehacer profesional de los docentes evaluados.

En virtud de los objetivos planteados para el presente estudio, la variable *Sexo* presentó DIF en el indicador 6.1: *Calidad del inicio de la clase*, en contra de los grupos focal “Hombres”, de acuerdo a los criterios utilizados para el análisis.

Por otro lado, para la variable *Zona Geográfica* no se observó indicadores con evidencia de DIF.

El análisis de la variable *Edad*, también arrojó evidencia de DIF en el indicador 6.1: *Calidad del inicio de la clase*, en contra del grupo focal correspondiente al rango etario de mayores de 61 años.

Al observar el comportamiento de los indicadores para el análisis de la variable *Nivel de Enseñanza*, es interesante como algunos indicadores mostraron un comportamiento similar, independiente del grupo considerado de referencia. De este modo, se observa que el indicador 1.2: *Relación entre actividades y objetivos* muestra consistentemente DIF para los grupos focales Segundo Ciclo Básico y Educación Media, tanto si se considera como grupo de referencia al grupo Primer Ciclo Básico, como el grupo Educación Especial. Lo mismo ocurre con el indicador 6.1 con respecto al grupo focal Educación Media, como con el indicador 6.2: *Calidad del cierre de la clase*, en contra del grupo focal Educación Parvularia.

Al analizar las variables *Nivel de Enseñanza* y *Asignatura*, es importante señalar que el indicador 7.4: *Indicador por asignatura o área*, es el único de los 20 indicadores que conforman este instrumento, que posee pautas de corrección diferenciadas específicamente para cada asignatura. Lo anterior es significativo al observar el comportamiento de este indicador y apreciar que presenta DIF severo contra el grupo de referencia “Educación Especial” en 3 de las 4 comparaciones, lo que implica que, pese a que los docentes pertenecientes a este grupo obtienen un desempeño global superior en el portafolio, se ven desfavorecidos frente a los grupos “Segundo Ciclo Básico”, “Educación Media” y Educación de Adultos”. Algo similar sucede al observar el comportamiento de este indicador en los análisis de la variable *Asignatura*, tanto en la comparación dentro del grupo “Segundo Ciclo” como, en menor medida, dentro del grupo “Educación Media”. Resultaría de especial interés, estudiar la influencia sobre los resultados que ejerce la posición relativa del indicador 7.4 dentro del instrumento y evaluar, al mismo tiempo, si existe la tendencia, por parte de los correctores, a compensar al final de la revisión a los docentes con bajo desempeño en los indicadores previos, esto puede, eventualmente, asomar como explicación a la varianza asociada a este proceso que expone Taut et al. (2011) y su relación con la variable asignatura.

En cuanto a la variable *Asignatura*, en el análisis correspondiente al segundo ciclo básico, es necesario señalar que todos los indicadores presentaron DIF severo al menos en una de las comparaciones. Los indicadores que presentaron menor frecuencia de DIF fueron 1.2; 4.1: *Atribución de los resultados de aprendizaje*; y 7.3: *Calidad de la retroalimentación a los alumno*, todos contra el grupo de referencia “Matemática”, y a favor de los grupos focales “Religión Católica”, “Inglés” y “Educación Física”, respectivamente.

Por contraparte, el indicador 6.1 presentó DIF severo contra 8 de los 9 grupos focales analizados, exceptuando únicamente al grupo “Lenguaje y Comunicación”, frente al cual presentó DIF moderado. El indicador 5.2: *Promoción de la participación de todos los alumnos*, también presenta DIF para gran parte de las comparaciones, en este caso no se observa funcionamiento diferencial únicamente contra los grupos “Educación Física” e “Inglés”, y se observa DIF contra el grupo de referencia para el grupo focal “Lenguaje y Comunicación”.

Esta misma variable, observada el grupo Educación Media muestra que el indicador 2.3: *Análisis de una clase*, es el único que no presenta DIF severo contra ningún grupo de comparación. Por otra parte, los indicadores 1.3: *Calidad de la secuencia entre las clases de la unidad* y 7.2: *Calidad de las preguntas*, presentan DIF contra el grupo de referencia en 7 de las comparaciones, lo que implica que el grupo “Matemática”, a pesar de obtener mejor desempeño general en el instrumento frente al resto de los grupos focales, se ve desfavorecido en la medición del atributo correspondiente a dichos indicadores.

Por el contrario, los indicadores 3.1: *Calidad de la evaluación* y 3.2: *Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada*, presentaron DIF en favor del grupo de referencia contra 7 grupos focales en cada caso, coincidiendo en “Artes Musicales”, “Artes Visuales”, “Educación Tecnológica”, “Lenguaje y

Comunicación”, “Inglés” y “Religión Católica”. También se puede observar una alta frecuencia de comparaciones con DIF en los indicadores 6.1 y 6.3: *Contribución de las actividades al logro de los objetivos de la clase*, mostrando funcionamiento diferencial contra 6 grupos focales cada indicador.

Resulta interesante la diferencia en la frecuencia de indicadores con DIF en relación al tipo de variable y la conformación de los grupos de comparación en cada una de ellas. La variable *Asignatura*, que presenta mayor cantidad de comparaciones identificadas con DIF, presenta, al mismo tiempo, la mayor cantidad de grupos de comparación, y una distribución altamente heterogénea en la cantidad de docentes que los conforman, esto último se replica en la variable *Nivel de Enseñanza*, evidenciando, igualmente, numerosos casos de DIF. De aquí se desprende la necesidad de revisar los procedimientos aplicados y la influencia de decisiones metodológicas relativas tres elementos cuyas relaciones parecen determinantes: la potencia estadística de los análisis, los tamaños muestrales recomendados (Cohen, 1988; S.-H. Kim et al., 2007; S. Kim et al., 1995; Zwick, 2012), y el número de grupos de comparación (Fidalgo & Mellenbergh, 1999; Fidalgo et al., 2000; S. Kim et al., 1995; Moses et al., 2013; Zwick, 2012; Zwick et al., 2013).

Finalmente, la presencia recurrente del indicador 6.1, con evidencia de funcionamiento diferencial no puede explicarse sin atender, al menos, a dos aspectos interrelacionados: 1) la definición del constructo y su operacionalización en este indicador específico, y 2) la estructura factorial del instrumento y su coherencia con la definición del constructo. Ambos aspectos podrían dar cuenta de deficiencias la construcción de este ítem que no satisfacen el proceso de aseguramiento de la validez del instrumento (Elosua, 2003; Gierl, 2005; Gómez-Benito & Hidalgo, 2005).

5.2 Conclusiones

Poniendo especial atención en las variables de tipo demográfico, se recomienda revisar la pauta que se aplica para la medición del indicador 6.1: *Calidad del inicio de la clase*, que presenta DIF severo, en la variable Sexo, contra el grupo focal *hombre*. Del mismo modo, este indicador presenta DIF en gran parte de las comparaciones realizadas, sin embargo, es necesario observar que, también en la mayoría de las comparaciones este indicador presenta impacto, por lo que las diferencias mostradas también se pueden presentar asociadas al nivel de habilidad del evaluado, es decir, el grupo de referencia respectivo obtiene mejor desempeño en este ítem, incluso cuando se controla por habilidad. En este sentido, pese a que existe cierta evidencia que analiza la interacción entre la presencia de impacto y el DIF, en estricto rigor, el análisis DIF no informa respecto de las causas del mismo, por lo que sus causas deben indagarse por medio de otros estudios.

Del mismo modo, frente a la presencia de DIF en variables relacionadas con aspectos de la formación, desarrollo y/o ejercicio profesional, se recomienda que estos resultados motiven estudios tendientes a encontrar las causas de este DIF. En este sentido, dentro de la variable *Nivel de Enseñanza*, resulta necesario revisar especialmente en los indicadores 1.2: *Relación entre actividades y objetivos* y 7.4: *Indicador por Asignatura*, indagando en variables como la dificultad del ítem, y las características específicas de los grupos encontrados en estas variables.

Los resultados de la variable *Asignatura*, sugieren una realidad heterogénea en relación con el desempeño frente a los diferentes indicadores. Estos resultados sugieren la necesidad de una profunda reflexión acerca de la pertinencia de que las diferencias propias de cada asignatura queden consideradas únicamente en el indicador 7.4. En relación con este indicador, es posible que su posición relativa, dentro del instrumento, afecte los resultados que presenta, ya que, puede existir por parte de los correctores, la tendencia a

compensar al final de la revisión a los docentes con bajo desempeño en los indicadores previos.

Cabe señalar que, en cierto sentido, la estandarización de los indicadores para hacer comparables los desempeños de docentes de diferentes áreas y niveles de enseñanza ha derivado, necesariamente, en obviar que aspectos específicos de cada asignatura puedan permear el desempeño observable en diversas competencias profesionales, por ejemplo aspectos como la naturaleza de las tareas asociadas a cada área, también elementos relacionados con la formación específica de cada disciplina, las herramientas didácticas y metodológicas de que se vale cada asignatura para el desarrollo de las habilidades específicas que ésta requiere. Por lo tanto, y a la luz de los resultados del presente estudio, se recomienda implementar análisis que permitan explicar las diferencias en el funcionamiento de los indicadores, indagando en variables propias de la caracterización de los diferentes grupos, como en elementos propios del instrumento tales como el nivel de dificultad de los indicadores, o inclusive, la posición de los mismos dentro de las pautas de corrección, y determinar, en virtud de ello, si las diferencias propias de estos grupos están suficientemente representadas en un único indicador.

Otro aspecto que requiere mayor investigación, y que excede los alcances del presente estudio, es la interacción entre Impacto y DIF en los indicadores del portafolio, dado que existe evidencia que señala una influencia en el incremento de la tasa de Error tipo I, para los análisis DIF por medios de análisis basados en tablas de contingencia, como los procedimientos MH (Aguerri et al., 2005).

Finalmente, y dadas las grandes implicancias asociadas a la aplicación de la Evaluación Docente, cuyas consecuencias no solo afectan a los docentes, quienes deben enfrentar este proceso de forma directa, sino también a todo el sistema educativo chileno. Es necesario que los instrumentos utilizados respondan de la forma más

satisfactoria posible a los criterios de rigor que puedan garantizar su justa implementación. En la vía de alcanzar los estándares de validez requeridos para este proceso evaluativo, se hace evidente la necesidad de complementar los datos emanados del presente trabajo y profundizar en estudios que aborden el funcionamiento diferencial, y potenciales sesgos, a través de aproximaciones metodológicas que puedan responder de forma más cabal a las características de este instrumento evaluativo, características tales como el tipo de respuesta (abierta), graduada en una escala ordenada, de naturaleza multidimensional, y aplicado a una población definida por una multiplicidad de grupos asociados a variables como Nivel de Enseñanza y/o Asignatura.

5.2.1 Limitaciones y Proyecciones

Es importante explicitar que existen algunas limitaciones para interpretar y generalizar los resultados del presente estudio. Por una parte, es necesario identificar elementos metodológicos que requieren atención, en primer lugar, los procedimientos utilizados para implementar los análisis consideran que el instrumento cumple con el supuesto de la unidimensionalidad, supuesto que permite determinar diferentes niveles de habilidad a partir de los puntajes obtenidos en la sumatoria de los ítems que conforman el instrumento. En la medida en que este instrumento no cumple con dicho supuesto, la dimensionalidad del portafolio puede constituirse en una limitación que reste validez a los resultados obtenidos.

Por otra parte, el instrumento portafolio tiene características que gran parte de los procedimientos que se han desarrollado para estudiar el DIF no considera, de modo que se hace mucho más difícil estimar el error asociado a adaptar dichos procedimientos. Por ejemplo, todos los procedimientos implementados de forma regular por el ETS corresponden a procedimientos desarrollados para evaluar el DIF en evaluaciones tipo prueba, por lo que no se involucra el error asociado al proceso de corrección. De acá se desprende que, para implementar estos procedimientos se debe asumir que el portafolio se comporta como una prueba con ítems politómicos, cuyo patrón de respuesta corresponde a una escala ordenada.

Otro aspecto que constituye una limitación en las posibilidades de análisis que se pueden alcanzar a partir de los resultados del presente trabajo, radica en la falta de acceso a las rúbricas de corrección aplicadas a los diferentes indicadores, ya que, al carecer de dicha información, no es posible determinar si las diferencias en los resultados de distintos grupos pudiera o no derivar de una conceptualización específica. Por ejemplo, si se observa el comportamiento del indicador 5.2 (*Promoción de la participación de todos los alumnos*) en el análisis de variable Asignatura, para la comparación de los grupos Matemática (referencia) y Artes Visuales (focal), se observa

que este indicador presenta tanto DIF severo, como Impacto, contra el grupo focal. Sin embargo, no es posible determinar si el comportamiento de este indicador se explica por la real diferencia en las competencias de los docentes a la hora de motivar la participación de los alumnos, o bien, si se debe a que el concepto de participación ha sido operacionalizado en la pauta de corrección, a partir de observaciones que no son pertinentes para ambas asignaturas.

Una tercera fuente de potenciales limitaciones se encuentra en aspectos coyunturales de la política nacional, y sus implicancias en las reformas educativas, de modo que, de acuerdo con las discusiones que precedieron la aprobación de la ley N°20903, se conformó una comisión técnica que actualmente se encuentra en proceso de elaboración de propuestas para mejorar la actual Evaluación Docente. Pese a que diversos organismos técnicos que participaron en la discusión parlamentaria señalaron la necesidad de desarrollar un sistema de evaluación que tome como base los instrumentos vigentes, según consta en el Primer Informe de la Comisión de Educación de la Cámara de Diputados (BCN, 2016), si esta voluntad no se materializa y el resultado de las propuestas deriva en un instrumento completamente nuevo, diferente del que ha sido objeto del presente estudio, los resultados arrojados por esta investigación tendrían escasa relevancia.

Sin embargo, y no obstante lo anterior, este trabajo presenta un potencial importante en dos ámbitos al menos, por un lado, la discusión acerca de las particularidades metodológicas constituye un acercamiento a la problemática de realizar este tipo de estudios sobre la base de instrumentos con las características del portafolio, de este modo, se puede considerar como un punto de partida para nuevos estudios, tanto en la búsqueda de potenciales fuentes de DIF identificado en algún ítem de este estudio, o bien como una referencia para nuevos estudios DIF que debieran implementarse en las futuras versiones de este instrumento. En este sentido, otro aspecto que requiere mayor investigación, y que excede los alcances del presente estudio, es la interacción entre

Impacto y DIF en los indicadores del portafolio, dado que existe evidencia que señala una influencia en el incremento de la tasa de Error tipo I, para los análisis DIF por medios de análisis basados en tablas de contingencia, como los procedimientos MH (Aguerri et al., 2005).

Por otra parte, considerando que el proceso de Evaluación Docente responde a una necesidad dinámica, cuyos instrumentos deben mantenerse en constante revisión, los resultados aportados acá pueden permitir tomar decisiones sobre la continuidad, o modificación, de elementos evaluados, o bien, pueden orientar la revisión de la forma en que se está abordando la medición de aspectos específicos.

En definitiva, la principal proyección y aspiración del presente trabajo es aportar en la discusión sobre el perfeccionamiento de los instrumentos que nos permitan mejorar la calidad de la educación en Chile.

REFERENCIAS

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*(1), 67–91. Retrieved from <http://www.jstor.org/stable/1434777>
- AERA, APA, & N. (1999). *NCME Standards for educational and psychological testing*. Washington, American Educational Research Association. Retrieved from [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:NCME+\(1999\)+Standards+for+educational+and+psychological+testing#0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:NCME+(1999)+Standards+for+educational+and+psychological+testing#0)
- Aguerri, M. E., Galibert, M. S., Zanelli, M. L., & Attorresi, H. F. (2005). Detección errónea del funcionamiento diferencial del ítem. Una comparación de métodos. *Psicothema, 17*(2), 350–355.
- Alvarado, M., & Cabezas, G. (2012). *La Evaluación Docente y sus instrumentos: discriminación del desempeño docente y asociación con los resultados de los estudiantes. MINEDUC-Programa de las Naciones Unidas para el Desarrollo*. Santiago, Chile. Retrieved from <http://www.divesup.cl/usuarios/bmineduc/doc/201206290959280.20120628WPEvaluacionDocenteFinal.pdf>
- Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 10*(2), 95–105. <https://doi.org/10.1111/j.1745-3984.1973.tb00787.x>
- Araya, C., Taut, S., Santelices, M. V., & Manzi, J. (2011). Validez consecucional del programa de asignación de excelencia pedagógica en Chile. *Estudios Pedagógicos (Valdivia), 37*(2), 25–42. <https://doi.org/10.4067/S0718-07052011000200001>
- Argüelles, M., Castillo, M. de los Á., & Saragoni, C. (2015). Sistema de evaluación e instrumentos de medición del Programa AEP. In B. Rodríguez, J. Manzi, C. Peirano, R. González, & D. Bravo (Eds.), *Reconociendo el mérito docente AEP 2002 - 2014* (pp. 24–39). Pontificia Universidad Católica de Chile.
- Avalos, B. (2002). *Profesores para Chile, historia de un proyecto*. Santiago, Chile.: Ministerio de Educación.
- Avalos, B., Cavada, P., Pardo, M., & Sotomayor, C. (2010). La profesión docente: temas y discusiones en la literatura internacional. *Estudios Pedagógicos (Valdivia), 36*(1). <https://doi.org/10.4067/S0718-07052010000100013>
- Baker, E., Barton, P., Haertel, E., Laad, H., Darling-Hammond, L., Linn, R. L., ... Shepard, L. (2010). Problems with the use of student test scores to evaluate teachers. *Economic Policy, 278*(278), 29. Retrieved from http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf
- Barber, M., & Mourshed, M. (2008). Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos. *Preal, (41)*, 1–48. Retrieved from <http://orton.catie.ac.cr/cgi->

bin/wxis.exe/?IsisScript=EARTH.xis&method=post&formato=2&cantidad=1&expresion=mfn=003423

- BCN. (2016). Historia de la Ley N° 20.903. Retrieved from www.bcn.cl/historiadelaLey
- Bellei, C., Contreras, D., & Valenzuela, J. P. (2008). Debate sobre la educación chilena y propuestas de cambio. In C. Bellei, D. Contreras, & J. P. Valenzuela (Eds.), *La agenda pendiente en educación* (pp. 13–46). Santiago, Chile.
- Bonifaz, R. (2011). Origen de la Evaluación Docente y su conexión con las políticas públicas en educación. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 13–32). Santiago, Chile.: MIDE UC.
- Braun, H., Chudowsky, N., & Koenig, J. (2010). *Getting value out of value added: Report of a workshop*.
- Bravo, D., Falck, D., González, R., Manzi, J., & Peirano, C. (2008). *La relación entre la evaluación docente y el rendimiento de los alumnos: Evidencia para el caso de Chile. ... de Chile. Proyecto ...* Santiago. Retrieved from http://redlaboral.microdatos.cl/docto_publicaciones/Evaluacion_docentes_rendimiento_escolar.pdf
- Burga, A. (2006). La unidimensionalidad de un instrumento de medición: perspectiva factorial. *Revista de Psicología de La Pontificia Universidad Católica Del Perú*, 24(1), 53–80.
- Camilli, G., & Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. SAGE Publications. Retrieved from <https://books.google.cl/books?id=kQOFIKjCILEC>
- Carlton, S. T., & Harris, A. M. (1992). Characteristics associated with differential item functioning on the scholastic aptitude test: Gender and majority/minority group comparisons. *ETS Research Report Series, 1992*(2), i-143. <https://doi.org/10.1002/j.2333-8504.1992.tb01495.x>
- Chang, H., & Mazzeo, J. (1996). Detecting DIF for polytomously scored items : An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333–353.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (Lawrence Erlbaum Associates, Ed.) (Segunda). New York, USA.: Hillsdale.
- Cohen, J. (1992). Quantitative methods in psychology. *Psychological Bulletin*.
- Cortés, F., & Lagos, M. J. (2011). Consecuencias de la Evaluación Docente. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 137–156). Santiago, Chile.: MIDE UC.
- Cuevas, M. L. (2013). *Sesgo cultural en los ítems de las pruebas del examen saber 11° en Colombia*. Universidad Nacional de Colombia. Retrieved from <http://www.bdigital.unal.edu.co/10773/>
- De Ayala, R. J., Kim, S.-H., Stapleton, L. M., & Dayton, C. M. (1999). *A reconceptualization of differential item functioning*. Montreal, Canada.

- Domínguez, S. (2012). Propuesta para el cálculo del alfa ordinal y theta de Armor. *Revista de Investigación En Psicología*, 15(1), 213–217.
- Domínguez, S. (2014). ¿Matrices policóricas/tetracóricas o matrices Pearson? Un estudio metodológico. *Revista Argentina de Ciencias Del Comportamiento*, 6(1), 39–48.
- Dorans, N. J. (2013). ETS Contributions to the quantitative assessment of item, test, and score fairness. *ETS Research Report Series*, 2013(2), i-38.
<https://doi.org/10.1002/j.2333-8504.2013.tb02334.x>
- Dorans, N. J., & Kulick, E. (1986). Demosnstrating the utility of the standarization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23(4), 355–368.
- Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. *Journal of Chemical Information and Modeling*. New Yersey.
- Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29(4), 309–319. Retrieved from <http://www.jstor.org/stable/1435087>
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7(2), 189–199.
<https://doi.org/10.1177/014662168300700207>
- Eisenberg, N. V. (2008). *The performance of teachers in Chilean public elementary schools: Exploring its relationship with teacher backgrounds and student achievement, and its distribution across schools and municipalities*. University of California. Retrieved from <http://search.proquest.com/docview/304661776>
- Elosua, P. (2003). Sobre la validez de los test. *Psichotema*, 15(2), 315–321.
- Elosua, P. (2006). Funcionamiento diferencial del ítem en la evaluación internacional PISA. Detección y comprensión. *RELIEVE. Revista Electrónica de Investigación Y Evaluación Educativa*, 12(2), 247–259. Retrieved from http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_4.htm
- Elosua, P., & López-Jáuregui, A. (2007). Aplicación de cuatro procedimientos de detección del funcionamiento diferencial sobre ítems politómicos. *Psicothema*, 19(2), 329–336.
- Elosua, P., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896–901.
- ETS. (2015). *2014 ETS standards for Quality and Fairness*. Retrieved from www.ets.org
- Fang, Z. (1996). A Review of Research on Teacher Beliefs and Practices.
- Ferreres Traver, D., González-Romá, V., & Gómez-Benito, J. (2000). Comparación del estadístico Mantel-Haenszel y la regresión logística en el funcionamiento diferencial de los ítems en dos pruebas de aptitud intelectual en un contexto bilingüe. *Psicothema*, 12(SUPPL. 2), 214–219.
- Feu i Gelis, J. (2004). La escuela rural en España: apuntes sobre las potencialidades

- pedagógicas, relacionales y humanas de la misma. *Revista Digital eRural, Educación, Cultura Y Desarrollo Rural*, (3). Retrieved from <http://educacin.upa.cl/revistaerural/erural.htm>
- Fidalgo, Á. (2005). Mantel-Haenszel Methods. In *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 1120–1126). John Wiley & Sons. Ltd. Chichester.
- Fidalgo, Á. (2011a). A new approach for differential item functioning detection using Mantel-Haenszel methods. The GMHDIF Program. *Spanish Journal of Psychology*, *14*(2), 1018–1022. https://doi.org/10.5209/rev_SJOP.2011.v14.n2.47
- Fidalgo, Á. (2011b). GMHDIF: A computer program for detecting DIF in dichotomous and polytomous items using generalized Mantel-Haenszel statistics. *Applied Psychological Measurement*, *35*(3), 247–249. <https://doi.org/10.1177/0146621610375691>
- Fidalgo, Á., & Bartram, D. (2010). A comparison between some generalized Mantel-Haenszel statistics for detecting DIF in data simulated under the graded response model. *Applied Psychological Measurement*, *34*(8), 600–606. <https://doi.org/10.1177/0146621610378405>
- Fidalgo, Á., & Ferreres Traver, D. (2002). Supuestos y consideraciones en los estudios empíricos sobre el funcionamiento diferencial de los ítems. *Psicothema*, *14*(2), 491–496.
- Fidalgo, Á., & Madeira, J. M. (2008). Generalized Mantel-Haenszel methods for differential item functioning detection. *Educational and Psychological Measurement*, *68*(6), 940–958. <https://doi.org/10.1177/0013164408315265>
- Fidalgo, Á., & Mellenbergh, G. J. (1999). Aplicación en una etapa, dos etapas e iterativamente de los estadísticos Mantel-Haenszel, 227–242.
- Fidalgo, Á., Mellenbergh, G. J., & Muñoz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, *5*(3), 43–53.
- Fidalgo, Á., Quintanilla, L., Fernández, R., Pons, F., & Aguerri, M. E. (2010). Detección del DIF en Ítems politómicos mediante el uso de los métodos Mantel-Haenszel. *Revista Electrónica de Metodología Aplicada*, *15*(1), 12–18.
- Fidalgo, Á., & Scalón, J. D. (2010). Using generalized Mantel-Haenszel statistics to assess DIF among multiple groups. *Journal of Psychoeducational Assessment*, *28*(1), 60–69. <https://doi.org/10.1177/0734282909337302>
- Flotts, M. P., & Andrea, A. (2011). El modelo de evaluación y los instrumentos. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 33–62). Santiago, Chile.: MIDE UC.
- Freiberg Hoffmann, A., Beatriz Stover, J., De la Iglesia, G., & Fernández Liporace, M. (2013). Correlaciones Policóricas y Tetracóricas En Estudios Factoriales Exploratorios y Confirmatorios. *Ciencias Psicológicas*, *VII*(2), 151–164.
- Gajardo, M., & De Andrea, A. M. (1992). *Docentes y docencia. Las zonas rurales*. Santiago, Chile.: UNESCO/FLACSO.

- García-huidobro, J. E. (2008). La Prueba Inicia : usos y abusos, (1991), 1–5.
- García, M. R. (2014). Informes técnicos MIDE UC/ Technical reports MIDE UC: Modelos de evaluación docente en el mundo e instrumentos utilizados para evaluar: ventajas , desventajas y consideraciones para su elección.
- Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice*, 24(1), 3–14. <https://doi.org/10.1111/j.1745-3992.2005.00002.x>
- Gómez-Benito, J., & Hidalgo, M. D. (1997). Evaluación del funcionamiento diferencial en ítems dicotómicos : una revisión metodológica. *Anuario de Psicología*, (74), 3–32.
- Gómez-Benito, J., & Hidalgo, M. D. (2003). Desarrollos Recientes En Psicometría, 1(1), 17–36.
- Gómez-Benito, J., & Hidalgo, M. D. (2005). La validez de los tests, escalas y cuestionarios. *La Sociología En Sus Escenarios*, 8(12), 1–14. Retrieved from <http://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/1750/1370>
- Gómez-Benito, J., Hidalgo, M. D., & Ferré, G. G. (2005). Introducción al funcionamiento diferencial del ítem. *Revista Psicología.com*, 9(2).
- Gómez-Benito, J., Hidalgo, M. D., & Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles Del Psicologo*, 31(1), 75–84.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The Market for Teacher Quality, 1–52. Retrieved from <http://www.nber.org/papers/w11154>
- Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Ite nls. *Applied Psychological Measurement*, 9(2), 139–164. <https://doi.org/10.1177/014662168500900204>
- Herrera, A. (2005). Detección de sesgo en los ítems mediante análisis de tablas de contingencia, 29–52.
- Hidalgo, M. D., & Gómez-Benito, J. (2000). Comparación de la eficacia de regresión logística polinómica y análisis discriminante logístico en la detección del DIF no uniforme. *Psicothema*, 12(SUPPL. 2), 298–300.
- Holland, P. W., & Thayer, D. T. (1986). *Differential Item Functioning and the Hantel-Haenszel Procedure*. Educational Testing Service. New Jersey.
- Holland, P. W., & Thayer, D. T. (1988). Differential Item Performance and the Mantel-Haenszel Procedure. In H. W. & H. I. Braun (Ed.), *Test Validity*. Routledge.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning. Differential item functioning*.
- Kim, S.-H., & Cohen, A. S. (1997). An investigation of the likelihood ratio test for detection of differential item functioning under the graded response model. Retrieved from <http://eric.ed.gov/?id=ED408304%5Cnhttp://files.eric.ed.gov/fulltext/ED408304.pdf>

- Kim, S.-H., Cohen, A. S., & Alagoz, C. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement, 44*(2), 93–116. Retrieved from 10.1111/j.1745-3984.2007.00029.x
- Kim, S., Cohen, A. S., & Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*(3), 261–276. Retrieved from http://www.jstor.org/stable/1435297?seq=1&cid=pdf-reference#references_tab_contents
- Landis, J. R., Heyman, E. R., & Koch, G. G. (1978). Average Partial Association in Three-Way Contingency Tables: A Review and Discussion of Alternative Tests. *International Statistical Review / Revue Internationale de Statistique, 46*(3), 237–254. <https://doi.org/10.2307/1402373>
- Lara, B., Mizala, A., & Repetto, A. (2010). Una Mirada a la Efectividad de los Profesores en Chile.
- León, M. G. (2008). *Calidad docente y rendimiento escolar en Chile*. Pontificia Universidad Católica de Chile.
- Ley N° 20903. Ley 20903: Crea el sistema de desarrollo profesional docente y modifica otras normas (2016). Chile: Biblioteca del Congreso Nacional de Chile. Retrieved from <https://www.leychile.cl/N?i=1087343&f=2016-04-01&p=>
- Linn, R. L., & Werts, C. E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement, 8*(1), 1–4. Retrieved from <http://www.jstor.org/stable/1434244>
- Longford, N. T. (2014). Screening Test Items for Differential Item Functioning. *Journal of Educational and Behavioral Statistics, 39*(1), 3–21. <https://doi.org/10.3102/1076998613515751>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems. The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference* (Vol. 1). Lawrence Erlbaum Associates Publishers.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88–91. <https://doi.org/10.3758/BF03192753>
- Mantel, N. (1963). Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel-Haenszel Procedure. *Journal of the American Statistical Association, 58*(February 2015), 690–700.
- Mantel, N., & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–748.
- Manzi, J., González, R., & Sun, Y. (2011). *La evaluación docente en Chile* (MIDE UC). Santiago: MIDE UC.
- Manzi, J., Meckes, L., Ortega, L., & Lacerna, P. (2011). ¿Qué características de la formación inicial de los docentes se asocian a mayores avances en su aprendizaje de conocimientos disciplinarios? Proyecto FONIDE N°: F511015, (DECEMBER), 1–155.

- Martínez-Rizo, F. (2012). Procedimientos para el estudio de las prácticas docentes. Revisión de la literatura. *RELIEVE-Revista Electrónica de ...*, 1–22. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&auth_type=crawler&jrnl=11344032&AN=77847564&h=3+czj1a+FqmewwxV1CeZD3zQ7L+EKdsBZDq59Z2zjj24PqCgVietobk06tBYg9QfjTP4bksCkE6TsXz2y/oeVw==&crl=c
- Martínez, J. F. (2013). Combinación de mediciones de la práctica y el desempeño docente: consideraciones técnicas y conceptuales para la evaluación docente. *Pensamiento Educativo: Revista de Investigación Educativa Latinoamericana*, 50(1), 4–20. <https://doi.org/10.7764/PEL.50.1.2013.2>
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational and Behavioral Statistics*, 7(2), 105–118. <https://doi.org/10.2307/1164960>
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107–122.
- MINEDUC. (2008). Marco para la buena enseñanza, 45.
- MINEDUC. (2013). *Boletín evaluación inicia*.
- MINEDUC. (2013). *Informe enfoque de género 2013*. Santiago, Chile.
- MINEDUC. (2013). *Resultados Evaluación Docente 2013*. Santiago, Chile.
- MINEDUC. (2014). *Orientaciones para la suscripción de convenios de desempeño colectivo*.
- Moses, T., Liu, J., Tan, A., Deng, W., & Dorans, N. J. (2013). Constructed-response DIF evaluations for mixed-format tests. *ETS Research Report Series, 2013(2)*, i-23. <https://doi.org/10.1002/j.2333-8504.2013.tb02340.x>
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and Simultaneous Item Bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315–328. <https://doi.org/10.1177/014662169401800403>
- OCDE. (2009). *Evaluación y reconocimiento de la calidad de los docentes*. Retrieved from www.oecd.org/publishing
- Oort, F. (1992). Using restricted factor analysis in test construction. Retrieved from <http://eric.ed.gov/?id=ED343959>
- Oshima, T. C., & Raju, N. S. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential. *Journal of Educational Measurement*, 34(3), 253. <https://doi.org/10.1111/j.1745-3984.1997.tb00518.x>
- Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19(1), 23–37. <https://doi.org/10.1177/014662169501900104>
- Prieto, M., & Contreras, G. (2008). Las concepciones que orientan las practicas evaluativas

- de los profesores: un problema a develar. *Estudios Pedagógicos (Valdivia)*, 34(2), 245–262. <https://doi.org/10.4067/S0718-07052008000200015>
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495–502.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14(2), 197–207. <https://doi.org/10.1177/014662169001400208>
- Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-Based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19(4), 353–368. <https://doi.org/10.1177/014662169501900405>
- Recio, P. (2012). *Equivalencia e invarianza de medida entre grupos: análisis factorial confirmatorio vs teoría de respuesta al ítem*. Programa de Doctorado en Psicología experimental y aplicada: Atención, aprendizaje y percepción. Universidad Complutense de Madrid. <https://doi.org/10.1073/pnas.0703993104>
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implication. *Journal of Education Statistics*, 4(3), 207–230.
- Rodríguez, B. (2015). Historia del programa y antecedentes generales. In B. Rodríguez, J. Manzi, C. Peirano, R. González, & D. Bravo (Eds.), *Reconociendo el mérito docente AEP 2002 - 2014* (pp. 10–23). Pontificia Universidad Católica de Chile.
- San Miguel, J. (2005). Nuevas demandas de la educación básica rural. *eRural*, 3(5). Retrieved from www.revistaerural.cl
- Santelices, M. V., Galleguillos, P., & González, J. (2015). Un estudio sobre la calidad docente en Chile : El rol del contexto en donde enseña el profesor y medidas de valor agregado, 24(11110262), 1–14. <https://doi.org/10.7764/psykhe.23.2.673>
- Santelices, M. V., & Valenzuela, F. (2015). Importancia de las características del profesor y de la escuela en la calidad docente: Una aproximación desde la Teoría de Respuesta del Ítem. *Estudios Pedagógicos (Valdivia)*, 41(2), 233–254. <https://doi.org/10.4067/S0718-07052015000200014>
- Santelices, M. V., & Wilson, M. (2010). Unfair treatment? The case of freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review*, 80(1), 106–133. <https://doi.org/10.1177/0020764011400999>
- Santiago, P., Benavides, F., Danielson, C., Goe, L., & Nusche, D. (2013). *OECD Reviews of Evaluation and Assessment in Education. Teacher Evaluation in Chile*. OECD. Retrieved from <http://dx.doi.org/10.1787/9789264172616-en>
- Schiefelbein, E. (1992). Relación entre calidad de la educación y el modelo de enseñanza frontal en América Latina. In *Boletín 29. PROYECTO PRINCIPAL DE EDUCACIÓN en América Latina y el Caribe* (pp. 3–18). Santiago, Chile.: UNESCO/OREALC.
- Schmitt, A. P., & Bleistein, C. A. (1987). Factors affecting differential item functioning for black examinees on scholastic aptitude test analogy items 1. *ETS Research Report Series*, 1987(1), i-46. <https://doi.org/10.1002/j.2330-8516.1987.tb00227.x>

- Shealy, R., & 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. <https://doi.org/10.1007/BF02294572>
- Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating Student Performance Measures into Teacher Evaluation Systems*. RAND. Retrieved from http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR917.pdf
- Sturman, M. C. (2003). Searching for the inverted U-Shaped relationship between time and performance: meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, *29*(6), 801–830. <https://doi.org/10.1016/S0149-2063>
- Sun, Y., Calderón, P., Valerio, N., & Torres, P. (2011). La implementación de la Evaluación Docente. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 61–90). Santiago, Chile.: MIDE UC.
- Sun, Y., Correa, M., Zapata, Á., & Carrasco, D. (2011). Resultados: qué dice la Evaluación Docente acerca de la enseñanza en Chile. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 91–136). Santiago, Chile.: MIDE UC.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Taut, S., Jiménez, D., & Barrientos, A. (2015). The quality of teacher evaluation systems: Validity evidence based on relations to other variables. COLMEE.
- Taut, S., Santelices, M. V., & Manzi, J. (2011). Estudios de validez de la Evaluación Docente. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 157–176). Santiago, Chile.: MIDE UC.
- Taut, S., Santelices, M. V., & Stecher, B. M. (2012). Validation of a national teacher assessment and improvement system. *Educational Assessment*, *17*(4), 163–199. <https://doi.org/10.1080/10627197.2012.735913>
- Taut, S., Santelices, M. V., & Valencia, E. (2010). *Informes Técnicos MIDE UC. Resultados de reevaluaciones y situación laboral de los docentes evaluados por el sistema de evaluación de desempeño docente entre 2003 y 2008*. Santiago, Chile.
- Taut, S., & Sun, Y. (2014). The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education Policy Analysis Archives*, *22*, N°71(Education), 33.
- Taut, S., Valencia, E., Palacios, D., Santelices, M. V., Jiménez, D., & Manzi, J. (2014). Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*. <https://doi.org/10.1080/0969594X.2014.961406>
- Thomas, C., & Hernández, R. (2005). El rol del profesor en la educación rural chilena. *Revista Digital eRural, Educación, Cultura Y Desarrollo Rural*, (3). Retrieved from

<http://www.revistaerural.cl/thyhe.htm>

- Toledo, G. A. (2010). Calidad docente y logro escolar: enfrentando el ordenamiento no aleatorio entre profesores y alumnos. *Segundo Congreso Interdisciplinario de Investigación En Educación 2012*.
- Valencia, E., & Manzi, J. (2011). Desempeño docente: relaciones con antecedentes de los profesores y su contexto. In J. Manzi, R. González, & Y. Sun (Eds.), *La Evaluación Docente en Chile* (pp. 177–194). Santiago, Chile.: MIDE UC.
- Vera, D., Osses, S., & Schiefelbein, E. (2012). Las Creencias de los profesores rurales: una tarea pendiente para la investigación educativa. *Estudios Pedagógicos (Valdivia)*, 38(1), 297–310. <https://doi.org/10.4067/S0718-07052012000100018>
- Villarroel, G., Acuña, M., De la Fuente, A., Fuentes, F., González, M., Gómez, M., ... Vidal, L. (2003). El Profesor Rural De Chiloe. *Revista Digital eRural, Educación, Cultura Y Desarrollo Rural*, 1–45.
- Williamson, G. (2004). Estudio sobre la educación para la población rural en Chile. In FAO - UNESCO - DGCS Italia - CIDE - REDUC (Ed.), *Educación para la población rural en Brasil, Chile, Colombia, Honduras, México, Paraguay y Perú*. (pp. 92–158). FAO.
- Wilson, S. M. (2009). Teacher Quality. Education Policy White Paper. *National Academy of Education (NJ1)*. Retrieved from <http://eric.ed.gov/?id=ED531145>
- Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857–860.
- Wu, L., & Tsai, R.-C. (2010). A comparison of three polytomous DIF detection methods.
- Zilberberg, A., Phan, H., Socha, A., Kong, J., & Keng, L. (2002). The effects of matching type and sample size on the Mantel-Haenszel technique for detecting items with DIF.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Electronic reprint of: Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29. <https://doi.org/10.1107/S0907444909031205>
- Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, 2012(1), i-30. <https://doi.org/10.1002/j.2333-8504.2012.tb02290.x>
- Zwick, R., & Thayer, D. T. (1996). Evaluating the magnitude of differential item functioning in polytomous items. *Journal of Educational and Behavioral Statistics*, 21(3), 187–201. Retrieved from <http://www.jstor.org/stable/1165267>
- Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Describing and categorizing DIF in polytomous items. *GRE Research*.
- Zwick, R., Ye, L., & Isham, S. (2013). An investigation of the efficacy of criterion refinement procedures in Mantel-Haenszel Dif analysis. *ETS Research Report Series*, 2013(1), i-23. <https://doi.org/10.1002/j.2333-8504.2013.tb02323.x>

ANEXOS

ANEXO 1: Análisis descriptivos e inferenciales

1 Análisis de valores perdidos

Tabla A1.1

Resumen de medias estimadas

	DOC_EDAD
Por lista	41.98
Todos los valores	41.97
EM	41.98
Regresión	41.98

Tabla A1.2

Resumen de desviaciones estándar estimadas

	DOC_EDAD
Por lista	11.551
Todos los valores	11.552
EM	11.551
Regresión	11.555

Tabla A.3

Correlación Promedio Portafolio/Suma Portafolio

		Prom_Pond_Portafolio	Suma_Portafolio
Prom_Pond_Portafolio	Correlación de Pearson	1	.986
	Sig. (bilateral)		.000
	Suma de cuadrados y productos vectoriales	1377.943	25863.140
	Covarianza	.086	1.605
	N	16114	16114
Suma_Portafolio	Correlación de Pearson	.986**	1
	Sig. (bilateral)	.000	
	Suma de cuadrados y productos vectoriales	25863.140	499670.971
	Covarianza	1.605	31.010
	N	16114	16114

** La correlación es significativa en el nivel 0,01 (2 colas).

2 Frecuencias por variables

Tabla A1.4

Frecuencia variable Sexo

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Masculino	4355	27.0	27.0	27.0
	Femenino	11759	73.0	73.0	100.0
	Total	16114	100.0	100.0	

Tabla A1.5

Frecuencias variable Zona Geográfica (Rural/Urbano)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Urbana	12169	75.5	75.5	75.5
	Rural	3945	24.5	24.5	100.0
	Total	16114	100.0	100.0	

Tabla A1.6

Frecuencias variable EDAD (por Rangos)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	30 o menos	3562	22.1	22.1	22.1
	31 a 35	2359	14.6	14.6	36.7
	36 a 40	1764	10.9	10.9	47.7
	41 a 45	1582	9.8	9.8	57.5
	46 a 50	2002	12.4	12.4	69.9
	51 a 55	2314	14.4	14.4	84.3
	56 a 60	2039	12.7	12.7	96.9
	61 o más	492	3.1	3.1	100.0
	Total	16114	100.0	100.0	

Tabla A1.7

Frecuencias variable Nivel de Enseñanza

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Educación Parvularia (EP)	1098	6.8	6.8	6.8
	Primer Ciclo Básico (PC)	4181	25.9	25.9	32.8
	Segundo Ciclo Básico (SC)	5303	32.9	32.9	65.7
	Educación Media (EM)	3286	20.4	20.4	86.1
	Educación Especial (EE)	1968	12.2	12.2	98.3
	Educación de Adultos (EA)	278	1.7	1.7	100.0
	Total	16114	100.0	100.0	

Tabla A1.8

Frecuencia variable Asignatura (Segundo Ciclo)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Artes Musicales	194	3.7	3.7	3.7
	Artes Visuales	176	3.3	3.3	7.0
	Ciencias Naturales	617	11.6	11.6	18.6
	Historia Geografía y Ciencias Sociales	591	11.1	11.1	29.8
	Educación Física	617	11.6	11.6	41.4
	Educación Tecnológica	93	1.8	1.8	43.1
	Inglés	503	9.5	9.5	52.6
	Lenguaje y Comunicación	1056	19.9	19.9	72.5
	Matemática	995	18.8	18.8	91.3
	Religión Católica	331	6.2	6.2	97.5
	Religión Evangélica	130	2.5	2.5	100.0
	Total	5303	100.0	100.0	

Tabla A.9

Frecuencias variable Asignatura (Educación Media)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Artes Musicales	193	5.9	5.9	5.9
	Artes Visuales	239	7.3	7.3	13.1
	Biología	179	5.4	5.4	18.6
	Educación Física	261	7.9	7.9	26.5
	Educación Tecnológica	167	5.1	5.1	31.6
	Física	95	2.9	2.9	34.5
	Filosofía y Psicología	138	4.2	4.2	38.7
	Historia, Geografía y Ciencias Sociales	395	12.0	12.0	50.7
	Inglés	496	15.1	15.1	65.8
	Lenguaje y Comunicación	394	12.0	12.0	77.8
	Matemática	425	12.9	12.9	90.7
	Química	119	3.6	3.6	94.4
	Religión Católica	134	4.1	4.1	98.4
	Religión Evangélica	51	1.6	1.6	100.0
	Total	3286	100.0	100.0	

3 Extracción de Componentes Principales

Tabla A1.10
Adecuación de la Matriz de Correlaciones Policóricas

Determinant of the matrix		0.092476101012932
Bartlett's statistic	Chi-cuadrado	38344.1
	Grados de libertad	190
	Sig.	0.000010
Kaiser-Meyer-Olkin (KMO) test		0.78339 (fair)

Tabla A1.11
Varianza explicada por el primer componente

Componente	Autovalor	Proporción de la varianza	Proporción de la varianza acumulada
1	3.90374	0.19519	0.19519
2	2.41527	0.12076	
3	1.42524	0.07126	
4	1.16953	0.05848	
5	1.08565	0.05428	
6	0.94598	0.0473	
7	0.8517	0.04258	
8	0.81321	0.04066	
9	0.78268	0.03913	
10	0.74592	0.0373	
11	0.73286	0.03664	
12	0.71173	0.03559	
13	0.65665	0.03283	
14	0.64383	0.03219	
15	0.62907	0.03145	
16	0.59592	0.0298	
17	0.56387	0.02819	
18	0.55314	0.02766	
19	0.49925	0.02496	
20	0.27477	0.01374	

4 Pruebas inferenciales por variables

4.1 Prueba T de Student variable Sexo

Tabla A1.12

Estadísticas de grupos. Variable Sexo

	Sexo	N	Media	Desviación estándar	Media de error estándar
Suma_Portafolio	Masculino	4355	45.15	5.847	.089
	Femenino	11759	46.79	5.395	.050
Prom_Pond_Portafolio	Masculino	4355	2.2369	.30567	.00463
	Femenino	11759	2.3305	.28324	.00261

Tabla A1.13

Prueba de muestras independientes. Variable Sexo

		Prueba de Levene de calidad de varianzas		prueba t para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Suma_Portafolio	Se asumen varianzas iguales	35.684	.000	-16.746	16112	.000	-1.640	.098	-1.832	-1.448
	No se asumen varianzas iguales			-16.140	7265.229	.000	-1.640	.102	-1.839	-1.441
Prom_Pond_Portafolio	Se asumen varianzas iguales	33.762	.000	-18.227	16112	.000	-.09360	.00513	-.10366	-.08353
	No se asumen varianzas iguales			-17.601	7290.384	.000	-.09360	.00532	-.10402	-.08317

4.2 Prueba T de Student variable Zona Geográfica

Tabla A1.14

Estadísticas de grupo. Variable Zona Geográfica

	Rural/Urbano	N	Media	Desviación estándar	Media de error estándar
Suma_Portafolio	Urbana	12169	46.47	5.612	.051
	Rural	3945	45.96	5.414	.086
Prom_Pond_Portafolio	Urbana	12169	2.3108	.29499	.00267
	Rural	3945	2.2878	.28373	.00452

Tabla A1.15

Prueba de muestras independientes

		Prueba de Levene de calidad de varianzas		prueba t para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Suma_Portafolio	Se asumen varianzas iguales	7.570	.006	5.051	16112	.000	.515	.102	.315	.715
	No se asumen varianzas iguales			5.145	6899.54	.000	.515	.100	.319	.711
Prom_Pond_Portafolio	Se asumen varianzas iguales	9.491	.002	4.305	16112	.000	.02305	.00535	.01256	.03355
	No se asumen varianzas iguales			4.391	6917.214	.000	.02305	.00525	.01276	.03334

4.3 ANOVA variable Edad

Tabla A1.16

Prueba de homogeneidad de varianzas

Puntaje Total Portafolio			
Estadístico de Levene	df1	df2	Sig.
.631	7	9992	.731

Tabla A1.17

ANOVA

Puntaje Total Portafolio					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	7102.130	7	1014.590	33.956	.000
Dentro de grupos	298552.347	9992	29.879		
Total	305654.478	9999			

4.3.1 Pruebas post hoc

Tabla A1.18

Comparaciones múltiples Rangos de Edad (36 a 40 años) / Variable dependiente: SUMA PORTAFOLIO

Prueba Bonferroni						
(I) Rangos de Edad	(J) Rangos de Edad	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
36 a 40	30 o menos	.285	.202	1.000	-.35	.92
	31 a 35	.111	.219	1.000	-.57	.79
	41 a 45	.765	.245	.050	.00	1.53
	46 a 50	.487	.228	.917	-.23	1.20
	51 a 55	.819	.221	.006	.13	1.51
	56 a 60	1.968	.225	.000	1.26	2.67
	61 o más	4.090	.355	.000	2.98	5.20

*. La diferencia de medias es significativa en el nivel 0.05.

4.4 ANOVA variable Nivel de Enseñanza

Tabla A1.19

Prueba de homogeneidad de varianzas

	Estadístico de	df1	df2	Sig.
	Levene			
Prom_Pond_Portafolio	30.057	5	16108	.000
Suma_Portafolio	32.014	5	16108	.000

Tabla A1.20

ANOVA

		Suma de	gl	Media	F	Sig.
		cuadrados		cuadrática		
Prom_Pond_Portafolio	Entre grupos	51.631	5	10.326	125.411	.000
	Dentro de grupos	1326.312	16108	.082		
	Total	1377.943	16113			
Suma_Portafolio	Entre grupos	16881.885	5	3376.377	112.651	.000
	Dentro de grupos	482789.086	16108	29.972		
	Total	499670.971	16113			

4.4.1 Pruebas post hoc

Tabla A1.21

Pruebas Tamhane, T3 Dunnett, Games-Howell, Dunnett C. Variable Nivel de Enseñanza (PC y EE)/Variable dependiente: Suma Portafolio

Prueba Tamhane		Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
(I) Nivel de Enseñanza	(J) Nivel de Enseñanza				Límite inferior	Límite superior
Primer Ciclo Básico (PC)	Educación Parvularia (EP)	3.226	.164	.000	2.74	3.71
	Segundo Ciclo Básico (SC)	1.787	.111	.000	1.46	2.11
	Educación Media (EM)	1.467	.127	.000	1.10	1.84
	Educación Especial (EE)	-.172	.142	.978	-.59	.24
	Educación de Adultos (EA)	2.742	.364	.000	1.67	3.82
Educación Especial (EE)	Educación Parvularia (EP)	3.398	.188	.000	2.85	3.95
	Primer Ciclo Básico (PC)	.172	.142	.978	-.24	.59
	Segundo Ciclo Básico (SC)	1.959	.144	.000	1.54	2.38
	Educación Media (EM)	1.639	.157	.000	1.18	2.10
	Educación de Adultos (EA)	2.915	.376	.000	1.81	4.02
Prueba T3 Dunnett						
Primer Ciclo Básico (PC)	Educación Parvularia (EP)	3.226	.164	.000	2.74	3.71
	Segundo Ciclo Básico (SC)	1.787	.111	.001	1.46	2.11
	Educación Media (EM)	1.467	.127	.000	1.10	1.84
	Educación Especial (EE)	-.172	.142	.978	-.59	.24
	Educación de Adultos (EA)	2.742	.364	.000	1.67	3.82
Educación Especial (EE)	Educación Parvularia (EP)	3.398	.188	.000	2.85	3.95
	Primer Ciclo Básico (PC)	.172	.142	.978	-.24	.59
	Segundo Ciclo Básico (SC)	1.959	.144	.000	1.54	2.38
	Educación Media (EM)	1.639	.157	.000	1.18	2.10
	Educación de Adultos (EA)	2.915	.376	.000	1.81	4.02
Prueba Games-Howell						
Primer Ciclo Básico (PC)	Educación Parvularia (EP)	3.226	.164	.000	2.76	3.69
	Segundo Ciclo Básico (SC)	1.787	.111	.000	1.47	2.10
	Educación Media (EM)	1.467	.127	.000	1.11	1.83
	Educación Especial (EE)	-.172	.142	.831	-.58	.23
	Educación de Adultos (EA)	2.742	.364	.000	1.70	3.79
Educación Especial (EE)	Educación Parvularia (EP)	3.398	.188	.000	2.86	3.93
	Primer Ciclo Básico (PC)	.172	.142	.831	-.23	.58
	Segundo Ciclo Básico (SC)	1.959	.144	.000	1.55	2.37
	Educación Media (EM)	1.639	.157	.000	1.19	2.09
	Educación de Adultos (EA)	2.915	.376	.000	1.84	3.99
Prueba Dunnett C						
Primer Ciclo Básico (PC)	Educación Parvularia (EP)	3.226	.164		2.76	3.69
	Segundo Ciclo Básico (SC)	1.787	.111		1.47	2.10
	Educación Media (EM)	1.467	.127		1.11	1.83
	Educación Especial (EE)	-.172	.142		-.58	.23
	Educación de Adultos (EA)	2.742	.364		1.70	3.79
Educación Especial (EE)	Educación Parvularia (EP)	3.398	.188		2.86	3.93
	Primer Ciclo Básico (PC)	.172	.142		-.23	.58
	Segundo Ciclo Básico (SC)	1.959	.144		1.55	2.37
	Educación Media (EM)	1.639	.157		1.19	2.09
	Educación de Adultos (EA)	2.915	.376		1.84	3.99

*. La diferencia de medias es significativa en el nivel 0.05.

4.5 ANOVA variable Asignatura (Segundo Ciclo Básico)

Tabla A1.22

Prueba de homogeneidad de varianzas

SUMA_PORTAFOLIO			
Estadístico de			
Levene	df1	df2	Sig.
2.026	10	5292	.027

Tabla A1.21

ANOVA

SUMA_PORTAFOLIO					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	20467.099	10	2046.710	68.100	.000
Dentro de grupos	159047.497	5292	30.054		
Total	179514.596	5302			

4.5.1 Pruebas post hoc

Tabla A1.23

Pruebas Tamhane, T3 Dunnett, Games-Howell y Dunnett C. Variable Asignatura-SC (Matemática)/Variable dependiente: Suma Portafolio

Prueba Tamhane		Diferencia de medias			95% de intervalo de confianza	
(I) Asignatura (SC)	(J) Asignatura (SC)	(I-J)	Error estándar	Sig.	Límite inferior	Límite superior
Matemática	Artes Musicales	2.322	.474	.000	.73	3.91
	Artes Visuales	5.198	.461	.000	3.65	6.74
	Ciencias Naturales	.533	.283	.966	-.41	1.47
	Historia Geografía y Ciencias Sociales	4.096	.279	.000	3.17	5.02
	Educación Física	4.287	.275	.000	3.37	5.20
	Educación Tecnológica	4.482	.588	.000	2.48	6.48
	Inglés	4.055	.310	.000	3.02	5.09
	Lenguaje y Comunicación	1.046	.231	.000	.28	1.81
	Religión Católica	5.507	.358	.000	4.32	6.70
	Religión Evangélica	4.146	.539	.000	2.33	5.97
Prueba T3 Dunnett						
Matemática	Artes Musicales	2.322	.474	.000	.73	3.91
	Artes Visuales	5.198	.461	.000	3.65	6.74
	Ciencias Naturales	.533	.283	.965	-.41	1.47
	Historia Geografía y Ciencias Sociales	4.096	.279	.000	3.17	5.02
	Educación Física	4.287	.275	.000	3.37	5.20
	Educación Tecnológica	4.482	.588	.000	2.48	6.48
	Inglés	4.055	.310	.000	3.02	5.09
	Lenguaje y Comunicación	1.046	.231	.000	.28	1.81
	Religión Católica	5.507	.358	.000	4.32	6.70
	Religión Evangélica	4.146	.539	.000	2.33	5.96
Prueba Games-Howell						
Matemática	Artes Musicales	2.322	.474	.000	.78	3.86
	Artes Visuales	5.198	.461	.000	3.70	6.70
	Ciencias Naturales	.533	.283	.728	-.38	1.45
	Historia Geografía y Ciencias Sociales	4.096	.279	.000	3.20	5.00
	Educación Física	4.287	.275	.000	3.40	5.17
	Educación Tecnológica	4.482	.588	.000	2.55	6.42
	Inglés	4.055	.310	.000	3.05	5.06
	Lenguaje y Comunicación	1.046	.231	.000	.30	1.79
	Religión Católica	5.507	.358	.000	4.35	6.66
	Religión Evangélica	4.146	.539	.000	2.38	5.91
Prueba Dunnett C						
Matemática	Artes Musicales	2.322	.474		.78	3.86
	Artes Visuales	5.198	.461		3.70	6.70
	Ciencias Naturales	.533	.283		-.38	1.45
	Historia Geografía y Ciencias Sociales	4.096	.279		3.20	5.00
	Educación Física	4.287	.275		3.40	5.18
	Educación Tecnológica	4.482	.588		2.54	6.42
	Inglés	4.055	.310		3.05	5.06
	Lenguaje y Comunicación	1.046	.231		.30	1.79
	Religión Católica	5.507	.358		4.35	6.67
	Religión Evangélica	4.146	.539		2.38	5.91

*. La diferencia de medias es significativa en el nivel 0.05.

4.6 ANOVA variable Asignatura (Educación Media)

Tabla A1.24

Prueba de homogeneidad de varianzas

SUMA_PORTAFOLIO			
Estadístico de Levene	df1	df2	Sig.
2.886	13	3272	.000

Tabla A1.25

ANOVA

SUMA_PORTAFOLIO					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	7643.028	13	587.925	18.762	.000
Dentro de grupos	102531.410	3272	31.336		
Total	110174.437	3285			

4.6.1 Pruebas post hoc

Tabla A1.26

Pruebas Tamhane y T3 Dunnett. Variable Asignatura (Matemática)/Variable dependiente: SUMA_PORTAFOLIO

Prueba Tamhane						
(I) Asignatura (EM)	(J) Asignatura (EM)	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
Matemática	Artes Musicales	3.251 [*]	.543	.000	1.36	5.14
	Artes Visuales	4.723 [*]	.474	.000	3.08	6.37
	Biología	3.581 [*]	.525	.000	1.75	5.41
	Educación Física	4.622 [*]	.444	.000	3.08	6.16
	Educación Tecnológica	3.365 [*]	.548	.000	1.45	5.28
	Física	1.780	.673	.568	-.60	4.16
	Filosofía y Psicología	1.132	.555	.981	-.81	3.07
	Historia, Geografía y Ciencias Sociales	3.009 [*]	.383	.000	1.68	4.34
	Inglés	2.988 [*]	.359	.000	1.75	4.23
	Lenguaje y Comunicación	1.487 [*]	.355	.003	.26	2.72
	Química	2.252 [*]	.554	.006	.31	4.20
	Religión Católica	5.480 [*]	.599	.000	3.38	7.58
	Religión Evangélica	4.670 [*]	.714	.000	2.08	7.26
Prueba T3 Dunnett						
(I) Asignatura (EM)	(J) Asignatura (EM)	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
Matemática	Artes Musicales	3.251 [*]	.543	.000	1.36	5.14
	Artes Visuales	4.723 [*]	.474	.000	3.08	6.37
	Biología	3.581 [*]	.525	.000	1.75	5.41
	Educación Física	4.622 [*]	.444	.000	3.08	6.16
	Educación Tecnológica	3.365 [*]	.548	.000	1.45	5.28
	Física	1.780	.673	.536	-.60	4.16
	Filosofía y Psicología	1.132	.555	.972	-.81	3.07
	Historia, Geografía y Ciencias Sociales	3.009 [*]	.383	.000	1.68	4.34
	Inglés	2.988 [*]	.359	.000	1.75	4.23
	Lenguaje y Comunicación	1.487 [*]	.355	.003	.26	2.72
	Química	2.252 [*]	.554	.006	.31	4.19
	Religión Católica	5.480 [*]	.599	.000	3.38	7.58
	Religión Evangélica	4.670 [*]	.714	.000	2.09	7.25

*. La diferencia de medias es significativa en el nivel 0.05.

Tabla A1.27

Pruebas Games-Howell y Dunnett C. Variable Asignatura (Matemática) /Variable dependiente: SUMA_PORTAFOLIO

Prueba Games-Howell

(I) Asignatura (EM)	(J) Asignatura (EM)	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
Matemática	Artes Musicales	3.251 [*]	.543	.000	1.41	5.09
	Artes Visuales	4.723 [*]	.474	.000	3.12	6.32
	Biología	3.581 [*]	.525	.000	1.81	5.35
	Educación Física	4.622 [*]	.444	.000	3.13	6.12
	Educación Tecnológica	3.365 [*]	.548	.000	1.51	5.22
	Física	1.780	.673	.323	-.52	4.08
	Filosofía y Psicología	1.132	.555	.741	-.75	3.01
	Historia, Geografía y Ciencias Sociales	3.009 [*]	.383	.000	1.72	4.30
	Inglés	2.988 [*]	.359	.000	1.78	4.20
	Lenguaje y Comunicación	1.487 [*]	.355	.003	.29	2.68
	Química	2.252 [*]	.554	.005	.37	4.14
	Religión Católica	5.480 [*]	.599	.000	3.45	7.51
	Religión Evangélica	4.670 [*]	.714	.000	2.18	7.16

Prueba Dunnett C

(I) Asignatura (EM)	(J) Asignatura (EM)	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
Matemática	Artes Musicales	3.251 [*]	.543		1.41	5.09
	Artes Visuales	4.723 [*]	.474		3.12	6.33
	Biología	3.581 [*]	.525		1.80	5.36
	Educación Física	4.622 [*]	.444		3.12	6.12
	Educación Tecnológica	3.365 [*]	.548		1.50	5.23
	Física	1.780	.673		-.53	4.09
	Filosofía y Psicología	1.132	.555		-.76	3.02
	Historia, Geografía y Ciencias Sociales	3.009 [*]	.383		1.72	4.30
	Inglés	2.988 [*]	.359		1.78	4.20
	Lenguaje y Comunicación	1.487 [*]	.355		.29	2.69
	Química	2.252 [*]	.554		.36	4.14
	Religión Católica	5.480 [*]	.599		3.44	7.52
	Religión Evangélica	4.670 [*]	.714		2.17	7.17

*. La diferencia de medias es significativa en el nivel 0.05.

ANEXO 2: Análisis de Impacto y DIF

1. ANÁLISIS IMPACTO Y DIF POR VARIABLE

1.1 SEXO

Tabla A2.1
Análisis Impacto y DIF variable Sexo

Indicadores	IMPACTO	DIF				SMD SMD/ σ_1
	D de Cohen	Análisis GMH N= 1000		Análisis GMH N= 10000		
	d	Q _{MH2}	p	Q _{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	0.03	7.93	0.005	45.424	0.000	0.12
1.2) Relación entre actividades y objetivos	-0.29	0.86	0.353	36.670	0.000	-0.16
1.3) Calidad de la secuencia entre las	-0.12	1.19	0.275	4.023	0.045	-0.01
2.1) Características de los alumnos consideradas en la unidad	-0.03	4.74	0.030	23.034	0.000	0.07
2.2) Análisis de la unidad realizada	-0.20	2.02	0.156	3.498	0.061	-0.10
2.3) Análisis de una clase	-0.03	0.47	0.494	12.926	0.000	0.07
3.1) Calidad de la evaluación	-0.17	0.91	0.340	4.023	0.045	-0.05
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.20	0.22	0.641	3.165	0.075	-0.07
4.1) Atribución de los resultados de aprendizaje	-0.20	0.94	0.332	3.248	0.072	-0.08
4.2) Calidad de la Retroalimentación	-0.18	0.01	0.908	3.721	0.054	-0.06
5.1) Promoción de la focalización de los alumnos en las actividades	-0.02	1.96	0.161	5.834	0.016	0.03
5.2) Promoción de la participación de todos los alumnos	-0.15	1.69	0.194	5.902	0.015	-0.07
5.3) Supervisión de las actividades	-0.16	0.43	0.514	11.768	0.001	-0.10
6.1) Calidad del inicio de la clase	-0.37	6.26	0.012	99.542	0.000	-0.28
6.2) Calidad del cierre de la clase	-0.19	0.56	0.454	7.281	0.007	-0.08
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.17	0.23	0.635	0.567	0.452	-0.06
7.1) Calidad de las explicaciones	0.02	8.80	0.003	23.600	0.000	0.12
7.2) Calidad de las preguntas	-0.05	0.25	0.618	3.051	0.081	0.04
7.3) Calidad de la retroalimentación a los alumnos	0.02	3.92	0.048	32.638	0.000	0.13
7.4) Indicador por asignatura o área	0.14	7.44	0.006	102.573	0.000	0.23

Nota: GR=Femenino v/s GF=Masculino

1.2 ZONA GEOGRÁFICA

Tabla A2.2
Análisis Impacto y DIF variable Zona Geográfica

Indicadores	IMPACTO		DIF			
	D de Cohen	Análisis GMH N= 1000		Análisis GMH N= 10000		SMD
		d	Q _{MH2}	p	Q _{MH2}	
1.1) Formulación y coherencia entre los objetivos	-0.01	0.00	0.98	1.88	0.17	-0.09
1.2) Relación entre actividades y objetivos	0.02	13.92	0.00	19.67	0.00	-0.08
1.3) Calidad de la secuencia entre las	-0.02	1.04	0.31	1.09	0.30	-0.11
2.1) Características de los alumnos consideradas en la unidad	-0.03	0.26	0.61	0.78	0.38	-0.11
2.2) Análisis de la unidad realizada	-0.08	9.37	0.00	7.50	0.01	-0.17
2.3) Análisis de una clase	-0.04	0.06	0.81	0.46	0.50	-0.13
3.1) Calidad de la evaluación	0.04	0.04	0.83	10.19	0.00	-0.05
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	0.01	1.50	0.22	2.79	0.09	-0.09
4.1) Atribución de los resultados de aprendizaje	-0.01	0.20	0.65	1.11	0.29	-0.11
4.2) Calidad de la Retroalimentación	-0.09	0.06	0.80	2.86	0.09	-0.19
5.1) Promoción de la focalización de los alumnos en las actividades	0.02	0.04	0.83	4.80	0.03	-0.03
5.2) Promoción de la participación de todos los alumnos	-0.05	0.33	0.57	2.48	0.12	-0.12
5.3) Supervisión de las actividades	0.01	1.31	0.25	4.14	4.14	-0.05
6.1) Calidad del inicio de la clase	0.02	0.25	0.62	3.38	3.38	-0.06
6.2) Calidad del cierre de la clase	-0.05	0.17	0.68	1.11	1.11	-0.14
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.08	1.78	0.18	3.61	3.61	-0.16
7.1) Calidad de las explicaciones	-0.11	7.84	0.01	14.26	14.26	-0.20
7.2) Calidad de las preguntas	-0.09	11.28	0.00	2.78	2.78	-0.17
7.3) Calidad de la retroalimentación a los alumnos	-0.08	10.07	0.00	3.93	3.93	-0.16
7.4) Indicador por asignatura o área	-0.09	9.58	0.00	6.82	6.82	-0.16

Nota: GR=Urbana v/s GF=Rural

1.3 EDAD

Tabla A2.3

Análisis Impacto y DIF variable Edad (por rangos), G3 vs G4

Indicadores	IMPACTO		DIF					
	D de Cohen	GMH Pareado	GMH Multigrupo				SMD	
			Global $\alpha=0.05$		Pareado $\alpha=0.0125$			
	d	Q _{MH2}	p	Q _{MH2}	p	Q _{MH2}	p	SMD/ σ_1
1.1) Formulación y coherencia entre los objetivos	-0.02	2.64	0.10	17.01	0.00	3.00	0.08	0.00
1.2) Relación entre actividades y objetivos	-0.05	1.25	0.26	2.03	0.73			-0.02
1.3) Calidad de la secuencia entre las	-0.04	1.41	0.24	5.55	0.24			-0.01
2.1) Características de los alumnos consideradas en la unidad	-0.05	0.07	0.79	7.37	0.12			-0.01
2.2) Análisis de la unidad realizada	-0.06	1.13	0.29	9.60	0.05	0.88	0.35	-0.03
2.3) Análisis de una clase	-0.01	0.09	0.77	4.69	0.32			0.02
3.1) Calidad de la evaluación	-0.04	2.43	0.12	6.88	0.14			-0.01
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.06	0.42	0.52	4.64	0.33			-0.03
4.1) Atribución de los resultados de aprendizaje	-0.04	0.33	0.56	1.40	0.84			-0.01
4.2) Calidad de la Retroalimentación	-0.01	0.02	0.90	4.08	0.40			0.03
5.1) Promoción de la focalización de los alumnos en las actividades	0.01	0.00	0.96	3.69	0.45			0.02
5.2) Promoción de la participación de todos los alumnos	-0.07	0.92	0.34	8.06	0.09			0.25
5.3) Supervisión de las actividades	0.00	0.16	0.68	16.09	0.00	0.12	0.72	0.02
6.1) Calidad del inicio de la clase	-0.04	0.00	0.99	11.35	0.02	0.00	0.95	-0.01
6.2) Calidad del cierre de la clase	-0.07	1.53	0.22	15.58	0.00	1.47	0.22	-0.03
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.07	1.10	0.29	6.78	0.15			-0.04
7.1) Calidad de las explicaciones	-0.04	0.96	0.33	5.38	0.25			0.00
7.2) Calidad de las preguntas	-0.01	2.49	0.11	7.19	0.13			0.03
7.3) Calidad de la retroalimentación a los alumnos	0.02	0.34	0.56	3.74	0.44			0.06
7.4) Indicador por asignatura o área	-0.03	0.32	0.57	13.58	0.01	0.57	0.45	0.00

Nota: G3=36 a 40 años (GR) vs G4=41 a 45 años (GF)

Tabla A2.4

Análisis Impacto y DIF variable Edad (por rangos), G3 vs G6

Indicadores	IMPACTO	DIF						SMD/σ _i
	D de Cohen	GMH Pareado		GMH Multigrupo				
	d	Q _{MH2}	p	Global α=0.05		Pareado α=0.0125		
		Q _{MH2}	p	Q _{MH2}	p	Q _{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	0.03	2.00	0.158	17.006	0.002	4.578	0.032	0.07
1.2) Relación entre actividades y objetivos	-0.03	0.22	0.641	2.027	0.731			0.03
1.3) Calidad de la secuencia entre las	-0.05	0.85	0.357	5.552	0.235			-0.01
2.1) Características de los alumnos consideradas en la unidad	0.00	1.68	0.195	7.374	0.117			0.04
2.2) Análisis de la unidad realizada	-0.08	3.33	0.068	9.599	0.048	0.785	0.376	-0.04
2.3) Análisis de una clase	0.02	0.22	0.637	4.692	0.320			0.06
3.1) Calidad de la evaluación	-0.05	0.82	0.364	6.882	0.142			-0.01
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.04	0.28	0.597	4.644	0.326			0.01
4.1) Atribución de los resultados de aprendizaje	-0.02	0.11	0.741	1.398	0.845			0.02
4.2) Calidad de la Retroalimentación	-0.02	0.83	0.362	4.078	0.396			0.03
5.1) Promoción de la focalización de los alumnos en las actividades	-0.09	0.18	0.672	3.691	0.449			-0.07
5.2) Promoción de la participación de todos los alumnos	-0.12	0.93	0.334	8.064	0.089			0.22
5.3) Supervisión de las actividades	-0.15	6.12	0.013	16.092	0.003	8.255	0.004	-0.13
6.1) Calidad del inicio de la clase	-0.12	0.08	0.780	11.349	0.023	1.627	0.202	-0.08
6.2) Calidad del cierre de la clase	-0.14	0.01	0.928	15.584	0.004	1.462	0.227	-0.10
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.12	1.18	0.278	6.776	0.148			-0.08
7.1) Calidad de las explicaciones	0.00	5.25	0.022	5.381	0.250			0.04
7.2) Calidad de las preguntas	-0.06	1.00	0.318	7.186	0.126			-0.01
7.3) Calidad de la retroalimentación a los alumnos	-0.05	1.33	0.248	3.739	0.443			0.00
7.4) Indicador por asignatura o área	0.01	3.77	0.052	13.578	0.009	0.047	0.828	0.05

Nota: G3=36 a 40 años (GR) vs G6=51 a 55 años (GF)

Tabla A2.5
Análisis Impacto y DIF variable Edad (por rangos), G3 vs G7

Indicadores	IMPACTO		DIF				SMD/σ _i		
	D de Cohen	GMH Pareado		GMH Multigrupo		SMD			
		Q _{MH2}	p	Q _{MH2}	p			Q _{MH2}	p
1.1) Formulación y coherencia entre los objetivos	-0.02	4.59	0.03	17.01	0.00	9.395	0.002	0.08	
1.2) Relación entre actividades y objetivos	-0.15	0.40	0.53	2.03	0.73			-0.02	
1.3) Calidad de la secuencia entre las	-0.07	0.05	0.83	5.55	0.24			0.03	
2.1) Características de los alumnos consideradas en la unidad	-0.07	1.12	0.29	7.37	0.12			0.03	
2.2) Análisis de la unidad realizada	-0.23	12.16	0.00	9.60	0.05	7.023	0.008	-0.12	
2.3) Análisis de una clase	-0.12	0.69	0.41	4.69	0.32			-0.01	
3.1) Calidad de la evaluación	-0.08	3.43	0.06	6.88	0.14			0.04	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.10	0.83	0.36	4.64	0.33			0.03	
4.1) Atribución de los resultados de aprendizaje	-0.13	0.17	0.68	1.40	0.84			-0.01	
4.2) Calidad de la Retroalimentación	-0.18	3.62	0.06	4.08	0.40			-0.07	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.11	0.01	0.91	3.69	0.45			-0.07	
5.2) Promoción de la participación de todos los alumnos	-0.12	0.53	0.47	8.06	0.09			0.37	
5.3) Supervisión de las actividades	-0.17	4.50	0.03	16.09	0.00	8.777	0.003	-0.11	
6.1) Calidad del inicio de la clase	-0.23	1.57	0.21	11.35	0.02	4.669	0.031	-0.14	
6.2) Calidad del cierre de la clase	-0.23	0.17	0.68	15.58	0.00	3.728	0.054	-0.12	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.11	0.20	0.65	6.78	0.15			-0.01	
7.1) Calidad de las explicaciones	-0.13	1.27	0.26	5.38	0.25			-0.02	
7.2) Calidad de las preguntas	-0.11	0.01	0.91	7.19	0.13			0.00	
7.3) Calidad de la retroalimentación a los alumnos	-0.11	0.36	0.55	3.74	0.44			0.01	
7.4) Indicador por asignatura o área	0.00	11.55	0.00	13.58	0.01	1.626	0.202	0.10	

Nota: G3=36 a 40 años (GR) vs G7=56 a 60 años (GF)

Tabla 2.6

Análisis Impacto y DIF variable Edad (por rangos), G3 vs G8

Indicadores	IMPACTO	DIF						SMD
	D de Cohen	GMH Pareado		GMH Multigrupo				
	d	Q _{MH2}	p	Global α=0.05		Pareado α=0.0125		
		Q _{MH2}	p	Q _{MH2}	p	Q _{MH2}	p	SMD/σ _i
1.1) Formulación y coherencia entre los objetivos	-0.12	8.68	0.00	17.01	0.00	11.981	0.001	0.12
1.2) Relación entre actividades y objetivos	-0.41	1.18	0.28	2.03	0.73			-0.08
1.3) Calidad de la secuencia entre las	-0.34	0.14	0.71	5.55	0.24			-0.07
2.1) Características de los alumnos consideradas en la unidad	-0.18	0.80	0.37	7.37	0.12			0.06
2.2) Análisis de la unidad realizada	-0.48	8.78	0.00	9.60	0.05	4.986	0.026	-0.19
2.3) Análisis de una clase	-0.31	3.85	0.05	4.69	0.32			-0.04
3.1) Calidad de la evaluación	-0.30	1.79	0.18	6.88	0.14			-0.01
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.32	2.50	0.11	4.64	0.33			0.02
4.1) Atribución de los resultados de aprendizaje	-0.36	0.96	0.33	1.40	0.84			-0.06
4.2) Calidad de la Retroalimentación	-0.37	4.95	0.03	4.08	0.40			-0.09
5.1) Promoción de la focalización de los alumnos en las actividades	-0.16	0.71	0.40	3.69	0.45			-0.06
5.2) Promoción de la participación de todos los alumnos	-0.36	3.83	0.05	8.06	0.09			0.53
5.3) Supervisión de las actividades	-0.34	8.77	0.00	16.09	0.00	7.883	0.005	-0.19
6.1) Calidad del inicio de la clase	-0.54	7.88	0.01	11.35	0.02	7.153	0.008	-0.30
6.2) Calidad del cierre de la clase	-0.41	4.64	0.03	15.58	0.00	4.751	0.029	-0.15
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.34	0.80	0.37	6.78	0.15			-0.06
7.1) Calidad de las explicaciones	-0.23	2.71	0.10	5.38	0.25			0.03
7.2) Calidad de las preguntas	-0.26	0.04	0.85	7.19	0.13			-0.01
7.3) Calidad de la retroalimentación a los alumnos	-0.22	1.52	0.22	3.74	0.44			0.04
7.4) Indicador por asignatura o área	-0.04	13.70	0.00	13.58	0.01	6.510	0.011	0.21

Nota: G3 = 36 a 40 años (GR) vs G8 = 61 años o más (GF)

1.4 NIVEL DE ENSEÑANZA

Tabla A2.7

Análisis Impacto y DIF variable Nivel de enseñanza, Primer Ciclo vs Educación Parvularia

Indicadores	IMPACTO		DIF					SMD/ σ_1
	D de Cohen	GMH Pareado	GMH Multigrupo				SMD	
			Global $\alpha=0.05$		Pareado $\alpha=0.0125$			
			Q_{MH2}	p	Q_{MH2}	p		
1.1) Formulación y coherencia entre los objetivos	-0.26	0.85	0.36	7.46	0.11			-0.07
1.2) Relación entre actividades y objetivos	-0.39	5.65	0.02	18.32	0.00	0.50	0.48	-0.10
1.3) Calidad de la secuencia entre las	-0.58	59.13	0.00	11.11	0.03	4.48	0.03	-0.36
2.1) Características de los alumnos consideradas en la unidad	-0.35	2.55	0.11	10.82	0.03	0.30	0.58	-0.12
2.2) Análisis de la unidad realizada	-0.16	7.12	0.01	8.60	0.07			0.09
2.3) Análisis de una clase	-0.32	3.52	0.06	4.55	0.34			-0.09
3.1) Calidad de la evaluación	-0.24	10.06	0.00	35.51	0.00	0.50	0.48	-0.01
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.29	7.07	0.01	16.33	0.00	0.01	0.91	-0.01
4.1) Atribución de los resultados de aprendizaje	-0.09	16.04	0.00	4.93	0.29			0.17
4.2) Calidad de la Retroalimentación	0.14	93.38	0.00	8.28	0.08			0.40
5.1) Promoción de la focalización de los alumnos en las actividades	0.22	51.68	0.00	7.84	0.10			0.34
5.2) Promoción de la participación de todos los alumnos	-0.18	1.70	0.19	2.73	0.60			0.53
5.3) Supervisión de las actividades	0.18	40.04	0.00	19.48	0.00	10.96	0.00	0.30
6.1) Calidad del inicio de la clase	0.17	52.46	0.00	30.60	0.00	3.12	0.08	0.42
6.2) Calidad del cierre de la clase	-0.61	78.12	0.00	16.06	0.00	8.88	0.00	-0.37
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.31	2.50	0.11	2.09	0.72			-0.07
7.1) Calidad de las explicaciones	-0.19	0.60	0.44	13.52	0.01	0.11	0.74	0.07
7.2) Calidad de las preguntas	0.39	126.52	0.00	18.36	0.00	10.01	0.00	0.58
7.3) Calidad de la retroalimentación a los alumnos	-0.85	190.20	0.00	38.64	0.00	26.11	0.00	-0.59
7.4) Indicador por asignatura o área	-0.62	92.53	0.00	15.29	0.00	9.84	0.00	-0.38

Nota: Primer Ciclo = Grupo de referencia; Educación Parvularia = Grupo Focal

Tabla A2.8

Análisis Impacto y DIF variable Nivel de enseñanza, Primer Ciclo vs Segundo Ciclo

Indicadores	IMPACTO		DIF					SMD/ σ_i	
	D de Cohen	d	GMH Pareado		GMH Multigrupo		SMD		
			Q_{MH2}	p	Global $\alpha=0.05$	Pareado $\alpha=0.0125$			
				Q_{MH2}	p	Q_{MH2}	p		
1.1) Formulación y coherencia entre los objetivos	0.04		39.90	0.00	7.46	0.11			0.13
1.2) Relación entre actividades y objetivos	-0.49		165.85	0.00	18.32	0.00	10.37	0.00	-0.34
1.3) Calidad de la secuencia entre las	-0.27		29.34	0.00	11.11	0.03	8.48	0.00	-0.16
2.1) Características de los alumnos consideradas en la unidad	-0.10		1.43	0.23	10.82	0.03	0.37	0.54	0.01
2.2) Análisis de la unidad realizada	-0.21		9.93	0.00	8.60	0.07			-0.09
2.3) Análisis de una clase	-0.04		8.50	0.00	4.55	0.34			0.07
3.1) Calidad de la evaluación	-0.45		150.62	0.00	35.51	0.00	27.28	0.00	-0.31
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.33		51.91	0.00	16.33	0.00	7.65	0.01	-0.18
4.1) Atribución de los resultados de aprendizaje	-0.20		5.10	0.02	4.93	0.29			-0.07
4.2) Calidad de la Retroalimentación	-0.08		0.66	0.42	8.28	0.08			0.05
5.1) Promoción de la focalización de los alumnos en las actividades	0.08		36.99	0.00	7.84	0.10			0.14
5.2) Promoción de la participación de todos los alumnos	-0.10		0.24	0.62	2.73	0.60			0.39
5.3) Supervisión de las actividades	-0.17		10.95	0.00	19.48	0.00	0.88	0.35	-0.10
6.1) Calidad del inicio de la clase	-0.23		18.35	0.00	30.60	0.00	3.52	0.06	-0.11
6.2) Calidad del cierre de la clase	-0.11		0.91	0.34	16.06	0.00	0.03	0.87	0.01
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.17		0.21	0.64	2.09	0.72			-0.05
7.1) Calidad de las explicaciones	0.11		78.70	0.00	13.52	0.01	4.06	0.04	0.23
7.2) Calidad de las preguntas	0.30		196.31	0.00	18.36	0.00	16.32	0.00	0.37
7.3) Calidad de la retroalimentación a los alumnos	0.11		59.10	0.00	38.64	0.00	1.78	0.18	0.24
7.4) Indicador por asignatura o área	-0.04		2.96	0.09	15.29	0.00	1.03	0.31	0.08

Nota: Primer Ciclo = Grupo de referencia; Segundo Ciclo = Grupo Focal

Tabla A2.9

Análisis Impacto y DIF variable Nivel de enseñanza, Primer Ciclo vs Educación Media

Indicadores	IMPACTO		DIF					SMD
	D de Cohen	GMH Pareado		GMH Multigrupo			SMD/ σ_1	
		Q _{MH2}	p	Global $\alpha=0.05$	Pareado $\alpha=0.0125$			
	d	Q _{MH2}	p	Q _{MH2}	p	Q _{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.05	8.15	0.00	7.46	0.11			0.02
1.2) Relación entre actividades y objetivos	-0.50	67.31	0.00	18.32	0.00	7.03	0.01	-0.38
1.3) Calidad de la secuencia entre las	-0.21	2.94	0.09	11.11	0.03	4.86	0.03	-0.13
2.1) Características de los alumnos consideradas en la unidad	0.10	39.69	0.00	10.82	0.03	0.74	0.39	0.18
2.2) Análisis de la unidad realizada	-0.02	17.25	0.00	8.60	0.07			0.08
2.3) Análisis de una clase	0.00	12.86	0.00	4.55	0.34			0.09
3.1) Calidad de la evaluación	-0.51	77.01	0.00	35.51	0.00	22.90	0.00	-0.40
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.39	40.99	0.00	16.33	0.00	10.61	0.00	-0.28
4.1) Atribución de los resultados de aprendizaje	-0.15			4.93	0.29			-0.05
4.2) Calidad de la Retroalimentación	-0.02	7.25	0.01	8.28	0.08			0.08
5.1) Promoción de la focalización de los alumnos en las actividades	0.06	13.63	0.00	7.84	0.10			0.11
5.2) Promoción de la participación de todos los alumnos	-0.17	7.26	0.01	2.73	0.60			0.28
5.3) Supervisión de las actividades	-0.22	17.46	0.00	19.48	0.00	1.61	0.20	-0.17
6.1) Calidad del inicio de la clase	-0.35	27.80	0.00	30.60	0.00	20.81	0.00	-0.26
6.2) Calidad del cierre de la clase	-0.17	0.04	0.84	16.06	0.00	6.88	0.01	-0.08
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.19	0.04	0.84	2.09	0.72			-0.09
7.1) Calidad de las explicaciones	0.26	127.66	0.00	13.52	0.01	8.42	0.00	0.36
7.2) Calidad de las preguntas	0.30	96.90	0.00	18.36	0.00	5.86	0.02	0.36
7.3) Calidad de la retroalimentación a los alumnos	-0.01	9.87	0.00	38.64	0.00	1.17	0.28	0.10
7.4) Indicador por asignatura o área	0.22	83.57	0.00	15.29	0.00	0.02	0.88	0.31

Nota: Primer Ciclo = Grupo de referencia; Educación Media = Grupo Focal

Tabla A2.10

Análisis Impacto y DIF variable Nivel de enseñanza, Primer Ciclo vs Educación de Adultos

Indicadores	IMPACTO		DIF					
	D de Cohen	GMH Pareado	GMH Multigrupo				SMD	
			Global $\alpha=0.05$		Pareado $\alpha=0.0125$			
	d	Q_{MH2}	p	Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i
1.1) Formulación y coherencia entre los objetivos	-0.01	3.47	0.06	7.46	0.11			0.14
1.2) Relación entre actividades y objetivos	-0.53	11.16	0.00	18.32	0.00	0.29	0.59	-0.32
1.3) Calidad de la secuencia entre las	-0.52	24.13	0.00	11.11	0.03	1.73	0.19	-0.34
2.1) Características de los alumnos consideradas en la unidad	0.12	31.83	0.00	10.82	0.03	13.53	0.00	0.16
2.2) Análisis de la unidad realizada	-0.16	1.02	0.31	8.60	0.07			0.32
2.3) Análisis de una clase	-0.26	1.48	0.22	4.55	0.34			0.04
3.1) Calidad de la evaluación	-0.42	3.35	0.07	35.51	0.00	0.93	0.34	-0.10
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.35	0.47	0.49	16.33	0.00	0.00	0.98	-0.18
4.1) Atribución de los resultados de aprendizaje	-0.17	3.04	0.08	4.93	0.29			-0.13
4.2) Calidad de la Retroalimentación	-0.26	0.33	0.57	8.28	0.08			0.04
5.1) Promoción de la focalización de los alumnos en las actividades	0.01	5.38	0.02	7.84	0.10			-0.12
5.2) Promoción de la participación de todos los alumnos	-0.28	5.53	0.02	2.73	0.60			-0.79
5.3) Supervisión de las actividades	-0.28	1.87	0.17	19.48	0.00	0.02	0.88	-0.24
6.1) Calidad del inicio de la clase	-0.45	12.50	0.00	30.60	0.00	0.07	0.79	-0.09
6.2) Calidad del cierre de la clase	-0.15	0.84	0.36	16.06	0.00	1.25	0.26	-0.21
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.33	0.74	0.39	2.09	0.72			0.04
7.1) Calidad de las explicaciones	0.07	12.83	0.00	13.52	0.01	2.08	0.15	-0.12
7.2) Calidad de las preguntas	0.18	17.51	0.00	18.36	0.00	4.05	0.04	0.40
7.3) Calidad de la retroalimentación a los alumnos	-0.16	0.43	0.51	38.64	0.00	0.00	0.98	0.04
7.4) Indicador por asignatura o área	-0.18	0.00	0.98	15.29	0.00	0.00	0.96	0.01

Nota: Primer Ciclo = Grupo de referencia; Educación de Adultos = Grupo Focal

Tabla A2.11

Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación Parvularia

Indicadores	IMPACTO	DIF						SMD/ σ_i
	D de Cohen	GMH Pareado		GMH Multigrupo				
	d	Q_{MH2}	p	Global $\alpha=0.05$		Pareado $\alpha=0.0125$		
		Q_{MH2}	p	Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.20	0.30	0.59	6.14	0.19			0.06
1.2) Relación entre actividades y objetivos	-0.59	0.24	0.62	44.73	0.00	2.58	0.11	-0.22
1.3) Calidad de la secuencia entre las	-0.19	0.91	0.34	5.19	0.27			0.09
2.1) Características de los alumnos consideradas en la unidad	-0.17	18.97	0.00	20.35	0.00	4.14	0.04	0.04
2.2) Análisis de la unidad realizada	-0.41	0.73	0.39	12.46	0.01	1.51	0.22	-0.16
2.3) Análisis de una clase	-0.41	6.06	0.01	7.90	0.10			-0.16
3.1) Calidad de la evaluación	-0.11	0.16	0.69	12.64	0.01	0.01	0.93	0.16
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.29	6.41	0.01	23.25	0.00	0.83	0.36	0.01
4.1) Atribución de los resultados de aprendizaje	-0.10	0.23	0.63	3.28	0.51			0.20
4.2) Calidad de la Retroalimentación	0.12	5.87	0.02	11.84	0.02	10.35	0.00	0.39
5.1) Promoción de la focalización de los alumnos en las actividades	-0.16	0.30	0.58	3.39	0.49			-0.06
5.2) Promoción de la participación de todos los alumnos	-0.28	6.14	0.01	5.19	0.27			0.38
5.3) Supervisión de las actividades	0.05	0.54	0.46	18.51	0.00	8.68	0.00	0.13
6.1) Calidad del inicio de la clase	0.05	0.97	0.32	23.73	0.00	4.24	0.04	0.23
6.2) Calidad del cierre de la clase	-0.82	0.04	0.85	25.34	0.00	13.75	0.00	-0.62
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.55	10.34	0.00	5.65	0.23			-0.39
7.1) Calidad de las explicaciones	-0.24	0.46	0.50	16.58	0.00	0.00	0.97	-0.01
7.2) Calidad de las preguntas	0.14	1.38	0.24	10.57	0.03	7.76	0.01	0.33
7.3) Calidad de la retroalimentación a los alumnos	-0.79	2.54	0.11	41.65	0.00	10.37	0.00	-0.55
7.4) Indicador por asignatura o área	0.09	7.28	0.01	102.95	0.00	11.72	0.00	0.25

Nota: Educación Especial = Grupo de referencia; Educación Parvularia = Grupo Focal

Tabla A2.12

Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Segundo Ciclo

Indicadores	IMPACTO		DIF					SMD/ σ_1
	D de Cohen	GMH Pareado	GMH Multigrupo				SMD	
			Global $\alpha=0.05$	Pareado $\alpha=0.0125$				
d	Q_{MH2}	p	Q_{MH2}	p	Q_{MH2}	p		
1.1) Formulación y coherencia entre los objetivos	0.08	30.07	0.00	6.14	0.19			0.20
1.2) Relación entre actividades y objetivos	-0.68	201.19	0.00	44.73	0.00	32.58	0.0000*	-0.49
1.3) Calidad de la secuencia entre las	0.05	22.53	0.00	5.19	0.27			0.21
2.1) Características de los alumnos consideradas en la unidad	0.06	23.52	0.00	20.35	0.00	9.44	0.00	0.17
2.2) Análisis de la unidad realizada	-0.47	89.60	0.00	12.46	0.01	11.38	0.00	-0.32
2.3) Análisis de una clase	-0.09	2.29	0.13	7.90	0.10			0.03
3.1) Calidad de la evaluación	-0.31	26.17	0.00	12.64	0.01	8.50	0.00	-0.15
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.34	28.06	0.00	23.25	0.00	13.58	0.00	-0.17
4.1) Atribución de los resultados de aprendizaje	-0.20	4.72	0.03	3.28	0.51			-0.04
4.2) Calidad de la Retroalimentación	-0.10	0.44	0.51	11.84	0.02	2.36	0.12	0.04
5.1) Promoción de la focalización de los alumnos en las actividades	-0.29	32.99	0.00	3.39	0.49			-0.24
5.2) Promoción de la participación de todos los alumnos	-0.20	9.26	0.00	5.19	0.27			0.26
5.3) Supervisión de las actividades	-0.29	58.14	0.00	18.51	0.00	0.23	0.63	-0.25
6.1) Calidad del inicio de la clase	-0.32	51.82	0.00	23.73	0.00	1.29	0.26	-0.24
6.2) Calidad del cierre de la clase	-0.34	45.89	0.00	25.34	0.00	5.30	0.02	-0.23
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.36	41.83	0.00	5.65	0.23			-0.29
7.1) Calidad de las explicaciones	0.09	24.80	0.00	16.58	0.00	8.67	0.00	0.19
7.2) Calidad de las preguntas	0.09	21.14	0.00	10.57	0.03	11.20	0.00	0.18
7.3) Calidad de la retroalimentación a los alumnos	0.13	40.38	0.00	41.65	0.00	11.19	0.00	0.26
7.4) Indicador por asignatura o área	0.68	432.38	0.00	102.95	0.00	79.49	0.00	0.76

Nota: Educación Especial = Grupo de referencia; Segundo Ciclo = Grupo Focal

TablaA2.13°

Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación Media

Indicadores	IMPACTO		DIF					SMD/ σ_i
	D de Cohen	Q _{MH2}	p	GMH		Q _{MH2}	p	
				Global $\alpha=0.05$	Multigrupo			
1.1) Formulación y coherencia entre los objetivos	0.00	10.47	0.00	6.14	0.19			0.10
1.2) Relación entre actividades y objetivos	-0.68	88.87	0.00	44.73	0.00	26.70	0.00	-0.53
1.3) Calidad de la secuencia entre las	0.11	30.81	0.00	5.19	0.27			0.24
2.1) Características de los alumnos consideradas en la unidad	0.25	68.31	0.00	20.35	0.00	9.60	0.00	0.34
2.2) Análisis de la unidad realizada	-0.28	4.17	0.04	12.46	0.01	3.33	0.07	-0.17
2.3) Análisis de una clase	-0.05	5.54	0.02	7.90	0.10			0.05
3.1) Calidad de la evaluación	-0.37	13.91	0.00	12.64	0.01	10.11	0.00	-0.24
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.40	22.42	0.00	23.25	0.00	18.90	0.00	-0.27
4.1) Atribución de los resultados de aprendizaje	-0.16			3.28	0.51			-0.03
4.2) Calidad de la Retroalimentación	-0.04	7.76	0.01	11.84	0.02	1.81	0.18	0.08
5.1) Promoción de la focalización de los alumnos en las actividades	-0.33	27.00	0.00	3.39	0.49			-0.28
5.2) Promoción de la participación de todos los alumnos	-0.26	12.11	0.00	5.19	0.27			0.16
5.3) Supervisión de las actividades	-0.33	39.83	0.00	18.51	0.00	1.36	0.24	-0.30
6.1) Calidad del inicio de la clase	-0.43	52.23	0.00	23.73	0.00	10.94	0.00	-0.37
6.2) Calidad del cierre de la clase	-0.39	31.73	0.00	25.34	0.00	16.11	0.00	-0.30
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.38	23.69	0.00	5.65	0.23			-0.32
7.1) Calidad de las explicaciones	0.24	80.01	0.00	16.58	0.00	12.24	0.00	0.33
7.2) Calidad de las preguntas	0.08	17.99	0.00	10.57	0.03	2.49	0.11	0.15
7.3) Calidad de la retroalimentación a los alumnos	0.02	14.27	0.00	41.65	0.00	1.19	0.27	0.12
7.4) Indicador por asignatura o área	0.88	428.35	0.00	102.95	0.00	60.15	0.00	0.95

Nota: Educación Especial = Grupo de referencia; Educación Media = Grupo Focal

Tabla A2.14

Análisis Impacto y DIF variable Nivel de enseñanza, Educación Especial vs Educación de Adultos

Indicadores	IMPACTO		DIF					SMD/ σ_i
	D de Cohen	GMH Pareado		GMH Multigrupo			SMD	
		Q_{MH2}	p	Global $\alpha=0.05$	Q_{MH2}	p		
1.1) Formulación y coherencia entre los objetivos	0.04	7.05	0.01	6.14	0.19			0.23
1.2) Relación entre actividades y objetivos	-0.69	21.21	0.00	44.73	0.00	2.95	0.09	-0.42
1.3) Calidad de la secuencia entre las	-0.12	0.20	0.66	5.19	0.27			0.10
2.1) Características de los alumnos consideradas en la unidad	0.27	49.92	0.00	20.35	0.00	15.95	0.00	0.42
2.2) Análisis de la unidad realizada	-0.41	4.59	0.03	12.46	0.01	0.03	0.85	-0.20
2.3) Análisis de una clase	-0.33	3.21	0.07	7.90	0.10			-0.15
3.1) Calidad de la evaluación	-0.27	0.54	0.46	12.64	0.01	1.05	0.31	-0.03
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.35	0.12	0.73	23.25	0.00	1.02	0.31	-0.10
4.1) Atribución de los resultados de aprendizaje	-0.17	4.13	0.04	3.28	0.51			0.07
4.2) Calidad de la Retroalimentación	-0.26	0.25	0.61	11.84	0.02	0.00	0.99	-0.05
5.1) Promoción de la focalización de los alumnos en las actividades	-0.36	7.21	0.01	3.39	0.49			-0.28
5.2) Promoción de la participación de todos los alumnos	-0.37	10.54	0.00	5.19	0.27			0.22
5.3) Supervisión de las actividades	-0.45	15.96	0.00	18.51	0.00	2.23	0.14	-0.38
6.1) Calidad del inicio de la clase	-0.54	23.16	0.00	23.73	0.00	0.04	0.84	-0.40
6.2) Calidad del cierre de la clase	-0.39	5.29	0.02	25.34	0.00	0.73	0.39	-0.22
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.59	15.17	0.00	5.65	0.23			-0.46
7.1) Calidad de las explicaciones	0.05	4.73	0.03	16.58	0.00	1.80	0.18	0.22
7.2) Calidad de las preguntas	-0.05	0.02	0.88	10.57	0.03	1.76	0.18	0.10
7.3) Calidad de la retroalimentación a los alumnos	-0.12	0.50	0.48	41.65	0.00	0.16	0.69	-0.10
7.4) Indicador por asignatura o área	0.50	60.90	0.00	102.95	0.00	18.99	0.00	0.61

Nota: Educación Especial = Grupo de referencia; Educación de Adultos = Grupo Focal

1.5 ASIGNATURA

1.5.1 Segundo Ciclo

Tabla A2.15

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Artes Musicales

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0055$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	0.08		12.91	0.00	10.38	0.001	0.25
1.2) Relación entre actividades y objetivos	-0.22		0.31	0.58			-0.04
1.3) Calidad de la secuencia entre las	0.28		41.57	0.00	35.66	0.000	0.47
2.1) Características de los alumnos consideradas en la unidad	-0.10		1.39	0.24			0.02
2.2) Análisis de la unidad realizada	0.09		13.53	0.00	11.82	0.001	0.24
2.3) Análisis de una clase	0.21		23.72	0.00	22.38	0.000	0.35
3.1) Calidad de la evaluación	-0.53		18.35	0.00	20.58	0.000	-0.37
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.49		10.98	0.00	12.79	0.000	-0.30
4.1) Atribución de los resultados de aprendizaje	-0.19		0.00	0.98			-0.02
4.2) Calidad de la Retroalimentación	0.11		15.95	0.00	13.88	0.000	0.27
5.1) Promoción de la focalización de los alumnos en las actividades	-0.07		0.07	0.79			0.00
5.2) Promoción de la participación de todos los alumnos	-0.43		11.91	0.00	14.35	0.000	-0.32
5.3) Supervisión de las actividades	-0.56		26.69	0.00	26.63	0.000	-0.51
6.1) Calidad del inicio de la clase	-1.77		207.05	0.00	216.75	0.000	-1.64
6.2) Calidad del cierre de la clase	-0.42		6.03	0.01	7.46	0.006	-0.24
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.17		0.56	0.46			-0.01
7.1) Calidad de las explicaciones	-0.29		3.29	0.07			-0.14
7.2) Calidad de las preguntas	0.00		3.65	0.06			0.15
7.3) Calidad de la retroalimentación a los alumnos	-0.38		6.53	0.01	8.24	0.000	-0.23
7.4) Indicador por asignatura o área	0.44		63.66	0.00	65.45	0.000	0.61

Nota: Matemática = Grupo de referencia; Artes Musicales = Grupo Focal

Tabla A2.16

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Artes Visuales

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0055$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.22	4.24	0.04	1.06	0.304	0.25	
1.2) Relación entre actividades y objetivos	-0.17	10.84	0.00	5.52	0.019	-0.04	
1.3) Calidad de la secuencia entre las	0.02	23.35	0.00	18.53	0.000	0.47	
2.1) Características de los alumnos consideradas en la unidad	0.04	20.07	0.00	14.33	0.000	0.02	
2.2) Análisis de la unidad realizada	-0.13	7.12	0.01	5.72	0.017	0.24	
2.3) Análisis de una clase	-0.01	16.16	0.00	12.54	0.000	0.35	
3.1) Calidad de la evaluación	-1.39	131.77	0.00	140.67	0.000	-0.37	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.24	88.04	0.00	103.66	0.000	-0.30	
4.1) Atribución de los resultados de aprendizaje	-0.40	0.12	0.73			-0.02	
4.2) Calidad de la Retroalimentación	-0.14	12.12	0.00	7.79	0.005	0.27	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.49	7.11	0.01	7.95	0.005	0.00	
5.2) Promoción de la participación de todos los alumnos	-1.05	65.92	0.00	67.63	0.000	-0.32	
5.3) Supervisión de las actividades	-0.50	9.97	0.00	9.71	0.002	-0.51	
6.1) Calidad del inicio de la clase	-1.05	47.25	0.00	43.51	0.000	-1.64	
6.2) Calidad del cierre de la clase	-0.42	1.36	0.24			-0.24	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.68	8.61	0.00	8.71	0.003	-0.01	
7.1) Calidad de las explicaciones	-0.16	7.73	0.01	10.72	0.001	-0.14	
7.2) Calidad de las preguntas	-0.36	1.27	0.26			0.15	
7.3) Calidad de la retroalimentación a los alumnos	-0.24	3.29	0.07			-0.23	
7.4) Indicador por asignatura o área	-0.06	15.84	0.00	18.68	0.000	0.61	

Nota: Matemática = Grupo de referencia; Artes Visuales = Grupo Focal

Tabla A2.17

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Historia, Geografía y Ciencias Sociales

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0055$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.08	12.95	0.00	13.73	0.000	0.22	
1.2) Relación entre actividades y objetivos	-0.17	8.03	0.00	8.21	0.004	0.19	
1.3) Calidad de la secuencia entre las	-0.71	48.53	0.00	43.51	0.000	-0.35	
2.1) Características de los alumnos consideradas en la unidad	-0.23	0.11	0.74			0.01	
2.2) Análisis de la unidad realizada	-0.13	9.27	0.00	10.41	0.001	0.17	
2.3) Análisis de una clase	-0.03	22.45	0.00	23.70	0.000	0.28	
3.1) Calidad de la evaluación	-0.28	1.67	0.20			0.03	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.19	8.16	0.00	8.86	0.003	0.15	
4.1) Atribución de los resultados de aprendizaje	-0.09	14.84	0.00	17.35	0.000	0.22	
4.2) Calidad de la Retroalimentación	-0.23	4.15	0.04	4.90	0.027	0.12	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.11	2.93	0.09			0.07	
5.2) Promoción de la participación de todos los alumnos	-0.69	63.53	0.00	57.26	0.000	-0.45	
5.3) Supervisión de las actividades	-0.28	5.68	0.02	3.84	0.050	-0.18	
6.1) Calidad del inicio de la clase	-1.08	185.52	0.00	173.02	0.000	-0.87	
6.2) Calidad del cierre de la clase	-0.38	2.98	0.08			-0.05	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.40	3.48	0.06			-0.15	
7.1) Calidad de las explicaciones	-0.72	45.51	0.00	38.69	0.000	-0.41	
7.2) Calidad de las preguntas	-0.39	0.76	0.38			-0.09	
7.3) Calidad de la retroalimentación a los alumnos	-0.30	3.13	0.08			-0.01	
7.4) Indicador por asignatura o área	-0.37	0.09	0.76			-0.01	

Nota: Matemática = Grupo de referencia; Historia, Geografía y Ciencias Sociales = Grupo Focal

Tabla A2.18

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Educación Física

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0055$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.06		21.42	0.00	21.20	0.000	0.27
1.2) Relación entre actividades y objetivos	-0.33		0.45	0.50			0.04
1.3) Calidad de la secuencia entre las	0.14		72.55	0.00	77.27	0.000	0.52
2.1) Características de los alumnos consideradas en la unidad	-0.44		10.04	0.00	9.23	0.002	-0.17
2.2) Análisis de la unidad realizada	-0.58		28.53	0.00	27.27	0.000	-0.27
2.3) Análisis de una clase	-0.39		5.11	0.02	7.36	0.007	-0.09
3.1) Calidad de la evaluación	-0.02		18.90	0.00	18.87	0.000	0.26
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.19		5.36	0.02	5.93	0.015	0.17
4.1) Atribución de los resultados de aprendizaje	-0.26		1.11	0.29			0.07
4.2) Calidad de la Retroalimentación	-0.05		20.41	0.00	26.58	0.000	0.28
5.1) Promoción de la focalización de los alumnos en las actividades	-0.04		17.78	0.00	15.81	0.000	0.11
5.2) Promoción de la participación de todos los alumnos	-0.29		0.01	0.94			-0.09
5.3) Supervisión de las actividades	-0.18		0.02	0.88			-0.09
6.1) Calidad del inicio de la clase	-1.50		333.19	0.00	349.70	0.000	-1.30
6.2) Calidad del cierre de la clase	-1.22		217.21	0.00	224.07	0.000	-0.91
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.20		4.93	0.03	5.02	0.025	0.08
7.1) Calidad de las explicaciones	-0.94		116.45	0.00	122.25	0.000	-0.68
7.2) Calidad de las preguntas	-0.23		0.12	0.73			0.09
7.3) Calidad de la retroalimentación a los alumnos	0.28		103.75	0.00	108.83	0.000	0.58
7.4) Indicador por asignatura o área	-0.16		5.19	0.02	6.95	0.008	0.22

Nota: Matemática = Grupo de referencia; Educación Física = Grupo Focal

Tabla A2.19

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Educación Tecnológica

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0055$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.60	7.96	0.00	6.60	0.010	-0.27	
1.2) Relación entre actividades y objetivos	-0.44	1.52	0.22			-0.08	
1.3) Calidad de la secuencia entre las	-0.28	0.29	0.59			0.08	
2.1) Características de los alumnos consideradas en la unidad	-0.04	5.61	0.02	7.79	0.005	0.21	
2.2) Análisis de la unidad realizada	-0.25	0.28	0.60			0.06	
2.3) Análisis de una clase	-0.21	0.47	0.49			0.08	
3.1) Calidad de la evaluación	-0.83	26.31	0.00	23.04	0.000	-0.52	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.97	38.62	0.00	34.68	0.000	-0.62	
4.1) Atribución de los resultados de aprendizaje	-0.37	0.11	0.74			-0.04	
4.2) Calidad de la Retroalimentación	-0.27	1.63	0.20			0.06	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.51	9.72	0.00	8.95	0.003	-0.37	
5.2) Promoción de la participación de todos los alumnos	-0.81	32.72	0.00	28.23	0.000	-0.59	
5.3) Supervisión de las actividades	-1.09	49.05	0.00	32.97	0.000	-1.01	
6.1) Calidad del inicio de la clase	-1.26	54.66	0.00	50.29	0.000	-1.01	
6.2) Calidad del cierre de la clase	-0.43	1.23	0.27			-0.10	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.29	0.58	0.45			0.00	
7.1) Calidad de las explicaciones	0.22	23.64	0.00	23.83	0.000	0.53	
7.2) Calidad de las preguntas	0.38	40.65	0.00	42.78	0.000	0.68	
7.3) Calidad de la retroalimentación a los alumnos	-0.18	0.29	0.59			0.11	
7.4) Indicador por asignatura o área	0.80	92.19	0.00	95.40	0.000	1.13	

Nota: Matemática = Grupo de referencia; Educación Tecnológica = Grupo Focal

Tabla A2.20

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Inglés

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0055$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.72		38.47	0.00	44.26	0.000	-0.42
1.2) Relación entre actividades y objetivos	-0.17		8.90	0.00	8.49	0.004	0.15
1.3) Calidad de la secuencia entre las	-0.29		0.00	0.96			0.04
2.1) Características de los alumnos consideradas en la unidad	0.05		34.58	0.00	37.69	0.000	0.27
2.2) Análisis de la unidad realizada	0.33		103.03	0.00	108.88	0.000	0.59
2.3) Análisis de una clase	-0.18		3.64	0.06			0.08
3.1) Calidad de la evaluación	-0.30		0.62	0.43			-0.01
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.19		8.77	0.00	8.70	0.003	0.14
4.1) Atribución de los resultados de aprendizaje	0.09		46.44	0.00	49.04	0.000	0.40
4.2) Calidad de la Retroalimentación	-0.16		9.98	0.00	10.59	0.001	0.15
5.1) Promoción de la focalización de los alumnos en las actividades	-0.16		1.21	0.27			-0.02
5.2) Promoción de la participación de todos los alumnos	-0.27		0.25	0.62			-0.07
5.3) Supervisión de las actividades	-0.25		1.96	0.16			-0.15
6.1) Calidad del inicio de la clase	-1.14		184.65	0.00	182.77	0.000	-0.92
6.2) Calidad del cierre de la clase	-0.39		3.34	0.07			-0.10
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.21		0.95	0.33			0.06
7.1) Calidad de las explicaciones	-1.00		128.78	0.00	133.52	0.000	-0.73
7.2) Calidad de las preguntas	-0.73		68.20	0.00	69.85	0.000	-0.46
7.3) Calidad de la retroalimentación a los alumnos	-0.46		7.17	0.01	9.22	0.002	-0.17
7.4) Indicador por asignatura o área	-0.54		14.13	0.00	14.39	0.000	-0.24

Nota: Matemática = Grupo de referencia; Inglés = Grupo Focal

Tabla A2.21

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Lenguaje y Comunicación

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0055$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.08	0.66	0.42			-0.01	
1.2) Relación entre actividades y objetivos	0.05	5.60	0.02	12.40	0.000	0.13	
1.3) Calidad de la secuencia entre las	-0.01	1.18	0.28			0.07	
2.1) Características de los alumnos consideradas en la unidad	-0.19	9.34	0.00	8.90	0.003	-0.13	
2.2) Análisis de la unidad realizada	-0.02	0.90	0.34			0.05	
2.3) Análisis de una clase	-0.02	0.75	0.39			0.04	
3.1) Calidad de la evaluación	-0.18	8.11	0.00	7.20	0.007	-0.11	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.35	36.12	0.00	31.87	0.000	-0.27	
4.1) Atribución de los resultados de aprendizaje	0.08	9.13	0.00	16.28	0.000	0.15	
4.2) Calidad de la Retroalimentación	0.09	9.42	0.00	16.97	0.000	0.17	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.20	15.80	0.00	14.30	0.000	-0.16	
5.2) Promoción de la participación de todos los alumnos	-0.45	77.39	0.00	73.24	0.000	-0.40	
5.3) Supervisión de las actividades	-0.12	5.37	0.02	3.77	0.052	-0.10	
6.1) Calidad del inicio de la clase	-0.27	24.22	0.00	23.20	0.000	-0.22	
6.2) Calidad del cierre de la clase	-0.36	42.74	0.00	40.41	0.000	-0.28	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.20	8.89	0.00	7.87	0.005	-0.14	
7.1) Calidad de las explicaciones	-0.31	33.16	0.00	30.44	0.000	-0.24	
7.2) Calidad de las preguntas	0.98	404.84	0.00	426.17	0.000	1.06	
7.3) Calidad de la retroalimentación a los alumnos	-0.22	13.57	0.00	11.70	0.001	-0.15	
7.4) Indicador por asignatura o área	0.07	5.29	0.02	7.52	0.006	0.17	

Nota: Matemática = Grupo de referencia; Lenguaje y Comunicación = Grupo Focal

Tabla NA2.22

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Religión Católica

Indicadores	IMPACTO		DIF			
	D de Cohen	d	GMH		SMD	
			$\alpha= 0.05$	$\alpha= 0.0055$		
		Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i
1.1) Formulación y coherencia entre los objetivos	-0.08	18.93	0.00	12.37	0.000	0.34
1.2) Relación entre actividades y objetivos	-0.14	21.02	0.00	16.97	0.000	0.32
1.3) Calidad de la secuencia entre las	-1.11	82.93	0.00	103.49	0.000	-0.64
2.1) Características de los alumnos consideradas en la unidad	-0.24	4.23	0.04	2.66	0.103	0.06
2.2) Análisis de la unidad realizada	-0.11	19.63	0.00	20.87	0.000	0.26
2.3) Análisis de una clase	-0.42	0.03	0.87			-0.05
3.1) Calidad de la evaluación	-0.52	6.83	0.01	8.72	0.003	-0.12
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.76	17.38	0.00	24.74	0.000	-0.29
4.1) Atribución de los resultados de aprendizaje	-0.51	0.52	0.47			-0.08
4.2) Calidad de la Retroalimentación	-0.33	5.66	0.02	3.10	0.078	0.09
5.1) Promoción de la focalización de los alumnos en las actividades	-0.19	1.45	0.23			0.00
5.2) Promoción de la participación de todos los alumnos	-0.70	34.75	0.00	44.97	0.000	-0.41
5.3) Supervisión de las actividades	-0.46	10.96	0.00	13.24	0.000	-0.34
6.1) Calidad del inicio de la clase	-1.06	75.43	0.00	84.74	0.000	-0.74
6.2) Calidad del cierre de la clase	-0.33	0.62	0.43			0.10
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.74	12.68	0.00	17.84	0.000	-0.41
7.1) Calidad de las explicaciones	0.05	50.14	0.00	40.71	0.000	0.42
7.2) Calidad de las preguntas	-0.60	15.90	0.00	26.06	0.000	-0.19
7.3) Calidad de la retroalimentación a los alumnos	-0.29	2.28	0.13			0.11
7.4) Indicador por asignatura o área	-0.22	10.60	0.00	5.23	0.022	0.21

Nota: Matemática = Grupo de referencia; Religión Católica = Grupo Focal

Tabla A2.23

Análisis Impacto y DIF variable Asignatura (Segundo Ciclo), Matemática vs Religión Evangélica

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0055$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.71		10.12	0.00	8.08	0.005	-0.43
1.2) Relación entre actividades y objetivos	-0.12		12.46	0.00	12.04	0.001	0.20
1.3) Calidad de la secuencia entre las	-0.07		11.36	0.00	11.70	0.001	0.26
2.1) Características de los alumnos consideradas en la unidad	0.10		19.34	0.00	20.64	0.000	0.33
2.2) Análisis de la unidad realizada	0.21		30.63	0.00	36.50	0.000	0.48
2.3) Análisis de una clase	-0.21		3.22	0.07			0.06
3.1) Calidad de la evaluación	-0.40		0.26	0.61			-0.12
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.43		0.07	0.79			-0.11
4.1) Atribución de los resultados de aprendizaje	-0.42		0.01	0.92			-0.13
4.2) Calidad de la Retroalimentación	-0.44		0.68	0.41			-0.13
5.1) Promoción de la focalización de los alumnos en las actividades	-0.30		0.69	0.41			-0.15
5.2) Promoción de la participación de todos los alumnos	-0.98		55.37	0.00	52.52	0.000	-0.77
5.3) Supervisión de las actividades	-0.34		2.06	0.15			-0.25
6.1) Calidad del inicio de la clase	-0.96		23.10	0.00	18.87	0.000	-0.74
6.2) Calidad del cierre de la clase	0.15		27.69	0.00	32.83	0.000	0.46
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.79		12.78	0.00	10.63	0.001	-0.54
7.1) Calidad de las explicaciones	-0.42		1.08	0.30			-0.13
7.2) Calidad de las preguntas	-0.54		5.10	0.02	5.10	0.024	-0.25
7.3) Calidad de la retroalimentación a los alumnos	-0.32		0.06	0.80			-0.04
7.4) Indicador por asignatura o área	0.13		29.56	0.00	36.52	0.000	0.46

Nota: Matemática = Grupo de referencia; Religión Evangélica = Grupo Focal

1.5.2 Educación Media

Tabla A2.24

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Artes Musicales

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha=0.05$		GMH $\alpha=0.0045$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.25		12.91	0.00	0.01	0.913	-0.02
1.2) Relación entre actividades y objetivos	-0.17		0.31	0.58			0.10
1.3) Calidad de la secuencia entre las	0.28		41.57	0.00	31.14	0.000	0.53
2.1) Características de los alumnos consideradas en la unidad	-0.11		1.39	0.24			0.08
2.2) Análisis de la unidad realizada	0.17		13.53	0.00	21.98	0.000	0.38
2.3) Análisis de una clase	-0.23		23.72	0.00	0.12	0.725	-0.04
3.1) Calidad de la evaluación	-0.54		18.35	0.00	15.41	0.000	-0.39
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.63		10.98	0.00	21.33	0.000	-0.39
4.1) Atribución de los resultados de aprendizaje	-0.33		0.00	0.98			-0.03
4.2) Calidad de la Retroalimentación	-0.06		15.95	0.00	6.14	0.013	0.18
5.1) Promoción de la focalización de los alumnos en las actividades	0.06		0.07	0.79			0.13
5.2) Promoción de la participación de todos los alumnos	-0.46		11.91	0.00	7.94	0.005	-0.32
5.3) Supervisión de las actividades	-0.37		26.69	0.00	4.95	0.026	-0.27
6.1) Calidad del inicio de la clase	-1.05		207.05	0.00	69.48	0.000	-0.89
6.2) Calidad del cierre de la clase	-0.41		6.03	0.01	6.09	0.014	-0.23
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.38		0.56	0.46			-0.19
7.1) Calidad de las explicaciones	-0.25		3.29	0.07			-0.05
7.2) Calidad de las preguntas	0.06		3.65	0.06			0.23
7.3) Calidad de la retroalimentación a los alumnos	-0.39		6.53	0.01	6.55	0.011	-0.17
7.4) Indicador por asignatura o área	0.37		63.66	0.00	46.92	0.000	0.58

Nota: Matemática = Grupo de referencia; Artes Musicales = Grupo Focal

Tabla A2.25

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Artes Visuales

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0045$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.54		3.04	0.08			-0.20
1.2) Relación entre actividades y objetivos	-0.24		4.68	0.03	3.58	0.059	0.16
1.3) Calidad de la secuencia entre las	0.25		50.57	0.00	48.37	0.000	0.58
2.1) Características de los alumnos consideradas en la unidad	0.13		29.25	0.00	28.87	0.000	0.42
2.2) Análisis de la unidad realizada	-0.09		9.89	0.00	7.95	0.005	0.22
2.3) Análisis de una clase	-0.14		4.40	0.04	4.40	0.036	0.18
3.1) Calidad de la evaluación	-1.73		157.62	0.00	167.71	0.000	-1.49
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.80		149.14	0.00	162.77	0.000	-1.43
4.1) Atribución de los resultados de aprendizaje	-0.44		2.10	0.15			-0.01
4.2) Calidad de la Retroalimentación	-0.10		14.01	0.00	13.52	0.000	0.26
5.1) Promoción de la focalización de los alumnos en las actividades	-0.32		0.64	0.43			-0.21
5.2) Promoción de la participación de todos los alumnos	-1.00		55.18	0.00	61.62	0.000	-0.80
5.3) Supervisión de las actividades	-0.50		9.58	0.00	11.89	0.001	-0.39
6.1) Calidad del inicio de la clase	-0.55		8.17	0.00	12.14	0.001	-0.34
6.2) Calidad del cierre de la clase	-0.33		0.11	0.74			-0.09
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.59		7.90	0.00	11.06	0.001	-0.35
7.1) Calidad de las explicaciones	0.34		66.05	0.00	58.00	0.000	0.62
7.2) Calidad de las preguntas	0.17		29.01	0.00	22.26	0.000	0.41
7.3) Calidad de la retroalimentación a los alumnos	-0.01		15.54	0.00	10.80	0.001	0.30
7.4) Indicador por asignatura o área	0.18		34.96	0.00	26.60	0.000	0.48

Nota: Matemática = Grupo de referencia; Artes Visuales = Grupo Focal

Tabla A2.26

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Biología

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0045$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.13		2.43	0.12			0.13
1.2) Relación entre actividades y objetivos	-0.73		21.17	0.00	18.33	0.000	-0.40
1.3) Calidad de la secuencia entre las	0.00		5.97	0.01	8.13	0.004	0.28
2.1) Características de los alumnos consideradas en la unidad	-0.44		3.69	0.05			-0.17
2.2) Análisis de la unidad realizada	-0.11		2.92	0.09			0.16
2.3) Análisis de una clase	-0.20		0.23	0.63			0.05
3.1) Calidad de la evaluación	-0.26		1.62	0.20			-0.08
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.78		26.07	0.00	22.13	0.000	-0.51
4.1) Atribución de los resultados de aprendizaje	0.13		24.06	0.00	31.04	0.000	0.48
4.2) Calidad de la Retroalimentación	-0.42		3.16	0.08			-0.10
5.1) Promoción de la focalización de los alumnos en las actividades	-0.23		1.39	0.24			-2.70
5.2) Promoción de la participación de todos los alumnos	-0.59		15.81	0.00	12.56	0.000	-0.41
5.3) Supervisión de las actividades	-0.46		4.94	0.03	4.96	0.026	-0.36
6.1) Calidad del inicio de la clase	-0.36		1.80	0.18			-0.17
6.2) Calidad del cierre de la clase	-0.35		0.85	0.36			-0.13
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.23		0.14	0.70			-0.01
7.1) Calidad de las explicaciones	0.41		59.07	0.00	65.18	0.000	0.65
7.2) Calidad de las preguntas	0.35		45.90	0.00	71.11	0.000	0.54
7.3) Calidad de la retroalimentación a los alumnos	-0.15		1.73	0.19			0.10
7.4) Indicador por asignatura o área	-0.46		4.44	0.04	2.45	0.117	-0.24

Nota: Matemática = Grupo de referencia; Biología = Grupo Focal

Tabla A2.27

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Educación Física

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH $\alpha= 0.05$		GMH $\alpha= 0.0045$		SMD
			Q_{MH2}	p	Q_{MH2}	p	
1.1) Formulación y coherencia entre los objetivos	-0.39		0.24	0.62			-0.03
1.2) Relación entre actividades y objetivos	-0.27		7.77	0.01	2.94	0.086	0.15
1.3) Calidad de la secuencia entre las	0.30		63.88	0.00	59.21	0.000	0.67
2.1) Características de los alumnos consideradas en la unidad	-0.45		2.55	0.11			-0.12
2.2) Análisis de la unidad realizada	-0.55		4.66	0.03	9.74	0.002	-0.21
2.3) Análisis de una clase	-0.46		0.93	0.34			-0.13
3.1) Calidad de la evaluación	-0.34		0.76	0.38			-0.11
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.50		0.00	0.95			-0.14
4.1) Atribución de los resultados de aprendizaje	-0.53		1.20	0.27			-0.08
4.2) Calidad de la Retroalimentación	-0.23		6.87	0.01	5.98	0.014	0.16
5.1) Promoción de la focalización de los alumnos en las actividades	0.19		12.87	0.00	11.16	0.001	0.34
5.2) Promoción de la participación de todos los alumnos	-0.20		0.17	0.68			0.00
5.3) Supervisión de las actividades	0.04		7.84	0.01	5.73	0.017	0.23
6.1) Calidad del inicio de la clase	-0.86		37.69	0.00	44.88	0.000	-0.65
6.2) Calidad del cierre de la clase	-1.06		72.32	0.00	72.60	0.000	-0.81
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.33		1.55	0.21			-0.03
7.1) Calidad de las explicaciones	-0.61		12.84	0.00	13.56	0.000	-0.33
7.2) Calidad de las preguntas	-0.09		2.45	0.12			0.20
7.3) Calidad de la retroalimentación a los alumnos	0.34		46.26	0.00	51.06	0.000	0.69
7.4) Indicador por asignatura o área	-0.06		2.77	0.10			0.28

Nota: Matemática = Grupo de referencia; Educación Física = Grupo Focal

Tabla A2.28

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Educación Tecnológica

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0045$	p	
		Q_{MH2}		Q_{MH2}		SMD/ σ_1	
1.1) Formulación y coherencia entre los objetivos	-0.79	29.17	0.00	30.01	0.000	-0.55	
1.2) Relación entre actividades y objetivos	-0.29	0.10	0.76			-0.01	
1.3) Calidad de la secuencia entre las	0.07	14.68	0.00	14.18	0.000	0.32	
2.1) Características de los alumnos consideradas en la unidad	0.00	7.90	0.00	7.69	0.006	0.21	
2.2) Análisis de la unidad realizada	-0.17	0.33	0.56			0.05	
2.3) Análisis de una clase	-0.18	0.12	0.73			0.03	
3.1) Calidad de la evaluación	-1.20	83.86	0.00	87.77	0.000	-1.03	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.45	117.80	0.00	122.05	0.000	-1.18	
4.1) Atribución de los resultados de aprendizaje	-0.29	0.00	0.96			0.01	
4.2) Calidad de la Retroalimentación	-0.40	1.57	0.21			-0.14	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.19	0.04	0.84			-0.10	
5.2) Promoción de la participación de todos los alumnos	-0.25	0.01	0.92			-0.11	
5.3) Supervisión de las actividades	-0.41	5.16	0.02	5.06	0.025	-0.29	
6.1) Calidad del inicio de la clase	-0.77	26.77	0.00	29.30	0.000	-0.62	
6.2) Calidad del cierre de la clase	-0.28	0.05	0.83			-0.10	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.44	2.86	0.09			-0.24	
7.1) Calidad de las explicaciones	0.58	73.57	0.00	66.44	0.000	0.79	
7.2) Calidad de las preguntas	0.60	85.45	0.00	81.03	0.000	0.79	
7.3) Calidad de la retroalimentación a los alumnos	-0.15	1.37	0.24			0.08	
7.4) Indicador por asignatura o área	1.46	233.38	0.00	226.84	0.000	1.66	

Nota: Matemática = Grupo de referencia; Educación Tecnológica = Grupo Focal

Tabla A2.29

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Historia, Geografía y Ciencias Sociales

Indicadores	IMPACTO		DIF			
	D de Cohen	GMH		GMH		SMD
		$\alpha=0.05$	$\alpha=0.0045$	$\alpha=0.0045$	$\alpha=0.0045$	
	d	Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_1
1.1) Formulación y coherencia entre los objetivos	-0.05	5.02	0.03	5.75	0.017	0.17
1.2) Relación entre actividades y objetivos	0.05	19.31	0.00	26.31	0.000	0.31
1.3) Calidad de la secuencia entre las	-0.64	19.40	0.00	25.16	0.000	-0.44
2.1) Características de los alumnos consideradas en la unidad	-0.07	9.09	0.00	9.58	0.002	0.14
2.2) Análisis de la unidad realizada	0.45	83.80	0.00	88.40	0.000	0.66
2.3) Análisis de una clase	-0.02	11.34	0.00	10.39	0.001	0.17
3.1) Calidad de la evaluación	-0.32	1.80	0.18			-0.19
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.45	5.46	0.02	5.98	0.015	-0.24
4.1) Atribución de los resultados de aprendizaje	-0.02	15.79	0.00	12.73	0.000	0.26
4.2) Calidad de la Retroalimentación	-0.33	0.01	0.91			-0.06
5.1) Promoción de la focalización de los alumnos en las actividades	-0.13	0.19	0.67			-0.05
5.2) Promoción de la participación de todos los alumnos	-0.65	24.07	0.00	34.20	0.000	-0.48
5.3) Supervisión de las actividades	-0.22	1.91	0.17			-0.13
6.1) Calidad del inicio de la clase	-0.50	12.70	0.00	14.22	0.000	-0.34
6.2) Calidad del cierre de la clase	-0.33	0.91	0.34			-0.13
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.66	23.73	0.00	31.52	0.000	-0.50
7.1) Calidad de las explicaciones	-0.50	5.96	0.01	13.08	0.000	-0.27
7.2) Calidad de las preguntas	0.09	32.88	0.00	24.54	0.000	0.26
7.3) Calidad de la retroalimentación a los alumnos	-0.28	1.49	0.22			-0.03
7.4) Indicador por asignatura o área	-0.32	0.02	0.90			-0.12

Nota: Matemática = Grupo de referencia; Historia, Geografía y Ciencias Sociales = Grupo Focal

Tabla A2.30

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Inglés

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0045$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.92	87.29	0.00	90.45	0.000	-0.26	
1.2) Relación entre actividades y objetivos	-0.20	0.14	0.70			0.02	
1.3) Calidad de la secuencia entre las	-0.13	2.29	0.13			0.53	
2.1) Características de los alumnos consideradas en la unidad	0.15	28.73	0.00	32.00	0.000	0.32	
2.2) Análisis de la unidad realizada	0.40	74.96	0.00	78.49	0.000	0.43	
2.3) Análisis de una clase	-0.17	0.16	0.69			0.10	
3.1) Calidad de la evaluación	-0.57	34.40	0.00	34.94	0.000	-0.46	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.42	10.82	0.00	11.29	0.001	-0.53	
4.1) Atribución de los resultados de aprendizaje	0.06	19.86	0.00	21.74	0.000	0.12	
4.2) Calidad de la Retroalimentación	-0.16	2.37	0.12			0.27	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.05	0.90	0.34			-0.14	
5.2) Promoción de la participación de todos los alumnos	-0.14	0.19	0.66			-0.42	
5.3) Supervisión de las actividades	0.04	2.68	0.10			-0.03	
6.1) Calidad del inicio de la clase	-0.57	36.85	0.00	33.93	0.000	-0.19	
6.2) Calidad del cierre de la clase	-0.32	4.88	0.03	4.31	0.038	-0.14	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.36	7.38	0.01	8.17	0.004	-0.42	
7.1) Calidad de las explicaciones	-0.58	31.17	0.00	29.64	0.000	-0.14	
7.2) Calidad de las preguntas	-0.25	1.42	0.23			1.20	
7.3) Calidad de la retroalimentación a los alumnos	-0.29	2.10	0.15			-0.51	
7.4) Indicador por asignatura o área	0.08	23.94	0.00	19.45	0.000	0.01	

Nota: Matemática = Grupo de referencia; Inglés = Grupo Focal

Tabla A2.31

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Lenguaje y Comunicación

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0045$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.42	22.13	0.00	17.34	0.000	-0.31	
1.2) Relación entre actividades y objetivos	-0.12	0.67	0.41			0.02	
1.3) Calidad de la secuencia entre las	0.48	49.22	0.00	69.40	0.000	0.58	
2.1) Características de los alumnos consideradas en la unidad	0.20	17.06	0.00	22.02	0.000	0.32	
2.2) Análisis de la unidad realizada	0.34	27.85	0.00	39.45	0.000	0.46	
2.3) Análisis de una clase	0.01	2.96	0.09			0.10	
3.1) Calidad de la evaluación	-0.53	26.10	0.00	33.02	0.000	-0.47	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.68	45.10	0.00	56.09	0.000	-0.59	
4.1) Atribución de los resultados de aprendizaje	0.00	1.71	0.19			0.13	
4.2) Calidad de la Retroalimentación	0.11	12.81	0.00	17.68	0.000	0.25	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.22	5.39	0.02	4.87	0.027	-0.17	
5.2) Promoción de la participación de todos los alumnos	-0.55	27.53	0.00	32.62	0.000	-0.46	
5.3) Supervisión de las actividades	-0.05	0.44	0.51			-0.02	
6.1) Calidad del inicio de la clase	-0.29	12.75	0.00	10.39	0.001	-0.22	
6.2) Calidad del cierre de la clase	-0.24	4.28	0.04	5.99	0.014	-0.15	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.46	26.79	0.00	26.04	0.000	-0.39	
7.1) Calidad de las explicaciones	-0.24	3.38	0.07			-0.14	
7.2) Calidad de las preguntas	1.02	168.71	0.00	205.04	0.000	1.11	
7.3) Calidad de la retroalimentación a los alumnos	-0.54	24.33	0.00	26.48	0.000	-0.43	
7.4) Indicador por asignatura o área	-0.12	0.31	0.58			0.01	

Nota: Matemática = Grupo de referencia; Lenguaje y Comunicación = Grupo Focal

Tabla A2.32

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Química

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0045$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.43	8.55	0.00	8.30	0.004	-0.28	
1.2) Relación entre actividades y objetivos	0.09	5.55	0.02	4.05	0.044	0.25	
1.3) Calidad de la secuencia entre las	-0.02	1.52	0.22			0.13	
2.1) Características de los alumnos consideradas en la unidad	-0.18	0.35	0.55			-0.04	
2.2) Análisis de la unidad realizada	0.13	4.63	0.03	3.55	0.060	0.28	
2.3) Análisis de una clase	-0.14	0.00	0.97			0.00	
3.1) Calidad de la evaluación	-0.44	7.81	0.01	9.75	0.002	-0.34	
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-0.09	0.58	0.45			0.05	
4.1) Atribución de los resultados de aprendizaje	0.03	4.53	0.03	2.61	0.106	0.22	
4.2) Calidad de la Retroalimentación	-0.15	0.08	0.78			0.02	
5.1) Promoción de la focalización de los alumnos en las actividades	-0.09	0.01	0.92			-0.04	
5.2) Promoción de la participación de todos los alumnos	-0.24	1.82	0.18			-0.13	
5.3) Supervisión de las actividades	-0.47	7.24	0.01	8.20	0.004	-0.39	
6.1) Calidad del inicio de la clase	-0.60	19.04	0.00	25.55	0.000	-0.49	
6.2) Calidad del cierre de la clase	0.13	4.42	0.04	4.42	0.036	0.26	
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.53	13.46	0.00	11.31	0.001	-0.41	
7.1) Calidad de las explicaciones	0.56	35.73	0.00	37.11	0.000	0.70	
7.2) Calidad de las preguntas	0.50	29.65	0.00	25.93	0.000	0.62	
7.3) Calidad de la retroalimentación a los alumnos	-0.52	13.67	0.00	15.84	0.000	-0.38	
7.4) Indicador por asignatura o área	-1.02	64.08	0.00	68.90	0.000	-0.89	

Nota: Matemática = Grupo de referencia; Química = Grupo Focal

Tabla A2.33

Análisis Impacto y DIF variable Asignatura (Educación Media), Matemática vs Religión Católica

Indicadores	IMPACTO		DIF				
	D de Cohen	d	GMH		GMH		SMD
			$\alpha= 0.05$	p	$\alpha= 0.0045$	p	
		Q_{MH2}	p	Q_{MH2}	p	SMD/ σ_i	
1.1) Formulación y coherencia entre los objetivos	-0.59		3.70	0.05			-0.20
1.2) Relación entre actividades y objetivos	0.12		33.27	0.00	35.37	0.000	0.60
1.3) Calidad de la secuencia entre las	-0.81		13.67	0.00	12.75	0.000	-0.40
2.1) Características de los alumnos consideradas en la unidad	-0.11		10.90	0.00	10.68	0.001	0.24
2.2) Análisis de la unidad realizada	0.11		28.76	0.00	27.38	0.000	0.48
2.3) Análisis de una clase	-0.39		0.29	0.59			-0.04
3.1) Calidad de la evaluación	-0.85		21.85	0.00	20.55	0.000	-0.57
3.2) Adecuación de la evaluación a los objetivos de aprendizaje de la unidad realizada	-1.11		37.14	0.00	40.28	0.000	-0.68
4.1) Atribución de los resultados de aprendizaje	-0.64		2.61	0.11			-0.14
4.2) Calidad de la Retroalimentación	-0.44		0.24	0.62			0.00
5.1) Promoción de la focalización de los alumnos en las actividades	-0.24		0.00	0.95			-0.08
5.2) Promoción de la participación de todos los alumnos	-0.57		4.28	0.04	3.51	0.061	-0.30
5.3) Supervisión de las actividades	-0.44		1.51	0.22			-0.26
6.1) Calidad del inicio de la clase	-0.64		2.26	0.13			-0.36
6.2) Calidad del cierre de la clase	-0.25		2.14	0.14			0.06
6.3) Contribución de las actividades al logro de los objetivos de la clase	-0.96		15.41	0.00	13.46	0.000	-0.67
7.1) Calidad de las explicaciones	0.41		58.66	0.00	66.07	0.000	0.76
7.2) Calidad de las preguntas	0.00		14.42	0.00	20.35	0.000	0.33
7.3) Calidad de la retroalimentación a los alumnos	-0.26		3.08	0.08			0.14
7.4) Indicador por asignatura o área	-0.30		3.64	0.06			0.06

Nota: Matemática = Grupo de referencia; Religión Católica = Grupo Focal