

Funcionamiento Diferencial del Ítem en una Evaluación Estandarizada según Necesidades Educativas Especiales Transitorias

Differential Item Functioning for Students with Special Educational Needs in a Primary Education Standardized Test

Fabiola Silva Morales y María Verónica Santelices Etchegaray*

Pontificia Universidad Católica de Chile

Un aspecto fundamental en la evaluación de aprendizajes es la validez de los resultados de la prueba utilizada, que no solo supone medir lo que se dice sino considerar las implicaciones de esos resultados, porque a partir de ellos se toman variadas decisiones. La existencia de funcionamiento diferencial del ítem (DIF) indica que sujetos que tienen igual grado de habilidad presentan diferente probabilidad de responder correctamente a un ítem, lo que se relaciona con el sesgo, es decir, favorecer a un grupo sobre otro en la evaluación. En los últimos años, la integración y el aprendizaje de los alumnos con necesidades educativas especiales transitorias (NEET) han sido el foco de políticas educacionales especiales que involucran recursos, especialistas y nuevas formas de trabajar en las escuelas de Chile. Mediante un enfoque metodológico secuencial cuantitativo se exploró el funcionamiento diferencial del ítem de una prueba estandarizada aplicada por un centro de medición en educación (MIDE UC) a más de 100 establecimientos en Chile en las áreas de lenguaje y matemática en los niveles desde 4° a 8° Básico para niños con y sin NEET. Un primer análisis cuantitativo exploró el funcionamiento diferencial del ítem (DIF) obteniéndose similitudes en la habilidad promedio de los grupos y pocos ítems con DIF significativo (menos del 10%). Posteriormente, se exploró la forma en que los colegios operacionalizan la variable integración, observándose que el grupo de niños sin NEET usado en el análisis estadístico puede haber sido, en realidad, un grupo híbrido que incluía a algunos niños con NEET.

Palabras clave: Evaluación, Sesgo, Funcionamiento diferencial del ítem (DIF), Sesgo estadístico, Integración educativa, Validez, Necesidades educativas especiales transitorias (NEET).

An important aspect of learning assessment is the validity of the results of the test used because they are the basis for social, political or institutional decisions. The existence of differential item functioning (DIF) indicates that subjects having presented the same degree of skill, have different probability of correctly respond to an item, which strongly affect an aspect of the validity of the test, which relates to the bias, i.e. favor one group over another in the evaluation. In recent years, integration and learning of students with temporary special educational needs (NEET) have been the focus of special education policies involving resources, specialists and new ways of working in schools in Chile. Using a sequential quantitative methodological approach we explored the differential item functioning, aspect of the validity of the results, of a standardized test administered by measurement center (MIDE UC) to more than 100 schools in Chile in language and mathematics from 4th to 8th grade to children with and without NEET. An initial analysis of differential item functioning (DIF) showed similarities in the average group ability and few items with significant DIF (less than 10%). Later, we explored how schools operationalize the variable depicting NEET. Results suggest that the group of children without NEET used in the statistical analysis may have been, in fact, a hybrid group including some children with NEET.

Keywords: Assessment, Bias, Differential item functioning (DIF), Statistical bias, Educational integration, Validity, Temporary special educational needs.

*Contacto: vsanteli@uc.cl

1. Introducción¹

La evaluación del aprendizaje es una herramienta de recolección de información respecto de los saberes de los alumnos tanto a nivel de aula como a gran escala. Dicha información determinará decisiones que involucran a variados agentes (Navas, 2012). La medición se debe realizar a través de instrumentos que aseguren criterios de calidad en las interpretaciones de los resultados obtenidos (Prieto y Delgado, 2003). La validez ya no es solo medir lo que se pretende, sino que, según Prieto y Delgado (2003), corresponde al grado en que las interpretaciones y los usos que se hacen de las puntuaciones están justificados científicamente, coincidiendo con los estándares internacionales para las pruebas educativas y psicológicas establecidos por la American Educational Research Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME). Es así como el proceso de validación de una prueba involucra la acumulación de evidencia para la interpretación y la relevancia de las puntuaciones de la prueba. Un aspecto de la validez de la prueba aplicada implica la exploración de la presencia de funcionamiento diferencial del ítem (DIF por su sigla en inglés) donde se distinguen dos grupos que son comparados, uno llamado de referencia (generalmente mayoritario) y el otro, focal, donde se sospecha que las propiedades psicométricas de los ítems puedan tener valores distintos (Gómez-Benito, Hidalgo y Guilera, 2010). La existencia de DIF es un antecedente relevante e indicativo de la posible existencia de sesgo. El sesgo es un “error sistemático originado por deficiencias en el test o en el modo en que éste es usado, que produce una distorsión en el significado de las puntuaciones y que adultera la interpretación propuesta” (Elosúa, 2003, p. 316). El sesgo representa un problema de equidad, ya que la puntuación obtenida por un sujeto en un test solo debe depender de su aprendizaje y no de sus características personales como etnia, género o necesidades educativas.

Desde 2010, Chile ha implementado políticas que promueven la inclusión de alumnos con necesidades educativas especiales permanentes (NEEP) y transitorias (NEET) proporcionando diversidad en las aulas. Sin embargo, la evaluación de aprendizajes que se realiza a nivel nacional no ha incorporado este aspecto en sus estándares de calidad.

El objetivo de esta investigación es analizar el DIF como un aspecto de la validez de una prueba estandarizada llamada prueba SEPA (Sistema de Evaluación para el Aprendizaje), aplicada en Chile por el centro de medición en evaluación de la Universidad Católica, MIDE UC, en las áreas de lenguaje y matemática desde 4° a 8° Básico, según la variable NEET. En ella, se aborda la interrogante respecto de si el puntaje obtenido en estas pruebas es reflejo del grado de habilidad que el alumno posee y si está libre de sesgo, es decir, es independiente de la existencia de necesidades educativas especiales transitorias (NEET).

Para complementar los resultados obtenidos, se exploró la forma en que los colegios que aplican la prueba SEPA reportan la variable NEET, indagando respecto de cómo se determina dicha clasificación, entre otros.

¹ Proyecto financiado en parte por el proyecto ANILLO SOC1107 de Conicyt Chile.

2. Fundamentación teórica

2.1. Inclusión

Durante las últimas décadas ha crecido la preocupación mundial por el derecho a la educación planteándose la necesidad de inclusión de personas con todo tipo de discapacidad en los sistemas educativos, generándose acuerdos internacionales en los que Chile participó (Organización de las Naciones Unidas, 2008). Es así como se genera un marco legal en torno a la definición y forma de trabajo de las NEET en los establecimientos del país. La inclusión escolar en Chile insta a cambiar el modo en que se concibe la discapacidad, desde una visión inicial centrada en la corrección de un defecto hacia una mirada donde la necesidad educativa es concebida como “resultado de la interacción entre los recursos, las carencias de los individuos y el medio” (Infante, Matus y Vizcarra, 2011) siendo parte del contexto existente, no arraigado en el déficit de la persona sino en las barreras sociales y culturales para evidenciar sus verdaderas capacidades (Tenorio, 2011).

De acuerdo con lo descrito, el diagnóstico de una NEET en establecimientos con financiamiento público en Chile es realizado por un equipo multidisciplinario de profesionales, que aplican una batería de pruebas estandarizadas y evalúan en varios ámbitos al estudiante, incluyendo lo cognitivo, afectivo-social, hábitos de estudio, familiar, salud y nutrición (Decreto 170/09). El marco legal permite que hasta el 20% de alumnos en aula presenten NEE, 7% permanentes y 13% transitorias. Si el número de alumnos por sala es en promedio 40, existirían 5 estudiantes con NEET y 2 con NEEP por aula. De no existir alumnos con NEEP, se permite la incorporación al proyecto de integración escolar (PIE) de un número máximo de 7 alumnos por aula.

El PIE es una estrategia inclusiva del sistema escolar chileno cuyo propósito es entregar apoyos adicionales a los estudiantes que presentan NEE permanentes o transitorias contribuyendo al mejoramiento de la calidad de la educación que se imparte en los establecimientos favoreciendo la presencia, la participación y el logro de los aprendizajes esperados (Decreto 170/09). A través del PIE, se entregan recursos a las escuelas para que los profesionales idóneos asistan pedagógicamente a los alumnos con necesidades educativas en el aula. Las escuelas, a su vez, se comprometen a promover una cultura que incentiva la inclusión y respeta la diversidad en el aula.

Según estadísticas del Ministerio de Educación (2013), un 6,3% de niños en Chile presentan NEE, concentrándose en el sector municipal y particular subvencionado. El 70% de ellos corresponde a NEET, las cuales se distribuyen en dificultades específicas de aprendizaje (DEA, 38%), trastorno específico del lenguaje (TEL, 21%), trastorno de déficit atencional (TDA, 18%) y coeficiente intelectual en rango limítrofe (FIL, 23%).

2.2. Adecuaciones

En las pruebas estandarizadas es posible que dadas las condiciones de aplicación del test se produzcan diferencias en los resultados que no sean propias de las diferencias en la habilidad. Para evitar esto, se realizan algunas adecuaciones que permitan a los estudiantes con NEET evidenciar su verdadero grado de aprendizaje (Sireci, Scarpati y Li, 2005). Varios investigadores han trabajado en esta línea, estudiando los efectos de la extensión del tiempo de aplicación del test, su presentación, el uso de ciertos materiales y equipos (calculadora) o la explicación en voz alta de algunos ítems y un ambiente de prueba tranquilo (Cahalan-Laitusis, 2010; Cawthon, 2010; Cohen, Gregg y Meng, 2005; Kettler et al., 2005; Mandinach, Bridgeman, Cahalan-Laitusis y Trapani, 2005).

En este sentido, Cohen et al. (2005) concluyen que las diferencias encontradas en el desempeño de alumnos con NEET y sin NEET no tienen su origen en la existencia de la necesidad educativa, sino que se explicarían por la falta de conocimiento del alumno, y que una extensión de tiempo solo beneficiaría al alumno con NEET que posee la habilidad matemática. Kettler et al. (2005) comparan el desempeño promedio entre niños sin NEET y con NEET encontrando que los alumnos con NEET presentan menores puntajes promedio sin adecuaciones y que, cuando se implementan, mejoran su desempeño.

Profundizando en el uso de adecuaciones, Mandinach et al. (2005) relacionan el impacto del tiempo extendido y la posibilidad de que la prueba se separe en secciones con el desempeño obtenido en el test SAT de lenguaje y matemática de alumnos con y sin NEET. Este estudio coincide con lo reportado por Cohen et al. (2005), planteando que las respuestas incorrectas se deben a la falta de conocimiento del alumno y no a la falta de tiempo. Sin embargo, contrasta con lo encontrado por Kettler et al. (2005), quienes verifican que la extensión de tiempo mejora los rendimientos en matemática más que en lenguaje, afectando principalmente a estudiantes con nivel de habilidad medio y/o alto, pero no habría ningún efecto cuando el estudiante posee bajo nivel de habilidad.

Mientras que algunos antecedentes muestran que los niños con NEET presentan menor rendimiento promedio en lenguaje y matemática (Kettler et al., 2005; Mandinach et al., 2005), Gregg (2007) comprueba que los estudiantes con NEET presentan rendimiento académico comparable a sus pares. En esta línea, Meltzer (2006) encuentra que los estudiantes con NEET solo requieren, en promedio, un semestre más que los alumnos sin NEET para cursar sus estudios superiores. Ello indicaría que no existen diferencias en la habilidad promedio entre los estudiantes con NEET y sin NEET.

Otros estudios exploran el menor rendimiento promedio de niños con NEET comprobando que estos presentan mayores índices de soledad y menores índices de autoeficacia académica (Idan y Margalit, 2014) y menor autorregulación en las tareas escolares y horas de estudio (Hen y Goroshit, 2014), lo que influiría en su rendimiento.

2.3. Estudios de DIF que involucran NEET

La mayoría de las investigaciones que relacionan DIF con NEET exploran el efecto de las adecuaciones sobre el desempeño demostrado en lenguaje y matemática. En ellas, el ajuste más frecuente y eficiente en lenguaje es extender el tiempo de aplicación. En matemática, en tanto, se encontró que algunos estudiantes con NEET mejoran su rendimiento cuando se realiza una explicación en voz alta de los ítems (Sireci et al., 2005).

Cho, Lee y Kingston en 2012 analizaron DIF según NEE en una prueba de matemática e historia aplicada a niños de 3° a 8° y encontraron un 21% de ítems con DIF con una tendencia mixta. Además, no hallaron diferencia en la habilidad promedio entre los grupos. Adicionalmente, el estudio muestra que algunos alumnos asignados con NEE respondían como alumnos sin NEE.

En 2013, Finch y Finch analizaron DIF en una prueba de lenguaje y matemática aplicada a alumnos de 3° de primaria en Estados Unidos, correlacionando multidimensiones (adecuación, escuela, género, etnia, raza, etc.) en un contexto de múltiples niveles. Los hallazgos reportados muestran que la existencia de DIF puede ser multicausal. Este argumento es compartido por el trabajo de Lin y Lin (2014)

quienes consideran que el DIF se explica mejor considerando la posible interacción entre variables tales como género, actitud hacia el aprendizaje, adecuación y habilidades.

En Latinoamérica existen pocos estudios que aborden DIF según NEET. Un aporte es la investigación de Moreira (2008), que comprueba la presencia de DIF en una prueba de matemática aplicada a alumnos de enseñanza media según trastorno de déficit atencional con hiperactividad (TDAH), encontrando que la prueba presenta un 18% de ítems con DIF de tamaño irrelevante pero estadísticamente significativo. Un análisis de los ítems con DIF indica que presentan vocabulario impreciso y redacción poco clara, entre otros, lo que perjudicaría a los niños con TDAH.

Dado que las políticas públicas mundiales en educación tienden a la inclusión de alumnos con NEET y que no conocemos de ningún estudio en Chile que explore DIF según esta variable en pruebas estandarizadas, resulta relevante estudiar este aspecto. Para ello se analizaron los resultados de una prueba aplicada a niños de 4° a 8° Básico en lenguaje y matemática en un total de más de 100 establecimientos educacionales; esta información se complementó con una encuesta que recopiló información general y específica relativa a las NEET.

3. Método

La investigación presenta un diseño secuencial, definido por Creswell (2008). En una primera etapa, se realizó un análisis estadístico para la determinación de los ítems con DIF seguido por un estudio complementario que indagó la forma en que se diagnostican, reportan y operacionalizan las NEET.

3.1. Análisis estadístico

3.1.1 Participantes

Alumnos de 4° a 8° Básico que respondieron la prueba SEPA de lenguaje y matemática aplicada por MIDE UC en colegios municipales y particulares subvencionados de Chile en 2012. La base de datos para el análisis se segmentó construyéndose muestras que varían entre 2452 y 2912 dependiendo del nivel de enseñanza y asignatura. Las muestras contienen alrededor de 3,4% de niños con NEE.

3.1.2. Instrumento

La prueba SEPA (Sistema de Evaluación Para el Aprendizaje) consiste en un conjunto de pruebas estandarizadas orientadas a evaluar los aprendizajes de los alumnos desde 1° Básico a III Medio en lenguaje y matemática a través de su trayectoria escolar (Pontificia Universidad Católica de Chile, 2014). Es confeccionada por el centro de medición MIDE UC, con base en el marco curricular nacional, y aporta información a docentes y directivos respecto del nivel de aprendizaje de los alumnos en el tiempo. Dicha prueba se aplica en colegios de todo Chile y dependiendo del nivel escolar posee 40 (desde 1° a 6° Básico) o 50 ítems (desde 7° en adelante) de selección múltiple. Durante su aplicación los colegios reportan los alumnos con integración, es decir, con NEE. Sin embargo, la información detallada respecto de cómo se diagnostica, categoriza y se evalúa a los alumnos que presentan NEE en los establecimientos se desconoce, puesto que los colegios presentan autonomía tanto para la declaración de la variable como para la implementación de adecuaciones durante su aplicación.

3.1.3. Procedimiento

Los datos de la prueba SEPA tomados en 2012 se segmentaron considerando: establecimientos que reportaron la variable integración, asignatura (lenguaje o matemática), nivel (4°, 5°, 6°, 7° y 8°) y dependencia (municipales y particulares subvencionados). El estudio consideró solo los niveles de 4° a 8° Básico de colegios municipales y particulares subvencionados que reportaron tener alumnos con NEET debido a que (1) los colegios particulares pagados presentan en promedio un mayor rendimiento que las otras dependencias y una baja presencia de niños con NEET, (2) los niños de 1° a 3° Básico podrían presentar dificultades en las habilidades lectoras propias de su etapa de desarrollo, y (3) el número de alumnos con NEET reportados por nivel disminuye en la medida que aumenta el nivel escolar.

3.1.4. Análisis de datos

La exploración del ajuste de los datos a un modelo unidimensional versus un modelo multidimensional, las diferencias promedio entre los niños con y sin NEET y la evaluación de ítems con DIF se realizaron utilizando el modelo de la Teoría de Respuesta al Ítem (TRI) de un parámetro, dado que los datos se ajustaron mejor a un modelo unidimensional. Este modelo se implementó para cada nivel (4° a 8° Básico) y para cada asignatura (lenguaje y matemática).

El modelo 1PL usado permite: 1) proporcionar mediciones de las variables educativas independientes del instrumento utilizado, y 2) disponer de instrumentos de medida cuyas propiedades no dependen de los objetos medidos (Muñiz, 1997, p. 18). Para Muñiz (1997), los dos supuestos involucrados en la TRI son la curva característica de los ítems (CCI) y la unidimensionalidad.

La CCI muestra que, a medida que aumenta el nivel de aptitud del individuo, se incrementa la probabilidad de acertar correctamente un ítem, lo que es descrito por una función matemática que relaciona la habilidad de los sujetos evaluados y la probabilidad de acertar correctamente un ítem. Dicha función queda definida por tres parámetros: dificultad (b), discriminación (a) y posibilidad de respuesta al azar (c) (Muñiz, 2010).

En un modelo 1PL, la probabilidad de respuesta correcta a un determinado ítem estará determinada por la habilidad del sujeto y la dificultad del ítem (Bond y Fox, 2001). La CCI para un modelo 1PL está dada por la ecuación (1), donde $P_i(\theta)$ es la probabilidad de que aleatoriamente un examinado con habilidad θ elija la respuesta correcta en el ítem i , b_i = parámetro dificultad del ítem i , I = número de ítems en el test y e = base del logaritmo natural cuyo valor aproximado es 2,718.

$$P_i(\theta) = e^{(\theta-b_i)} / [1 + e^{(\theta-b_i)}] \quad I = 1, 2, \dots, I \quad (1)$$

Para el análisis del ajuste de los datos al modelo, la determinación de las diferencias entre grupos y el análisis de ítems con DIF, se utilizó el software Conquest. La forma general del modelo de respuesta al ítem provisto por Conquest es el Modelo Logístico Multinomial Multidimensional de Coeficientes Aleatorios (MRCML) descrito por Adams, Wilson y Wang (1997). El MRCML es una extensión de los modelos que utilizan la TRI de tipo Rasch o 1PL, donde la probabilidad de un sujeto de responder correctamente es función de su capacidad y la dificultad de los ítems. Conquest estima la dificultad de los ítems controlando por el efecto de la variable NEET y el nivel de habilidad (Adams et al., 1997; Wu, Adams y Wilson, 1998).

3.1.5. Estadísticos para evaluar el ajuste del modelo, la diferencia en habilidad entre grupos y la significancia estadística de DIF

Para el análisis de las pruebas se exploró, en primer lugar, el grado de ajuste de los datos al modelo unidimensional versus un modelo multidimensional de cinco dimensiones en lenguaje y cuatro en matemática definidos de acuerdo a las tablas de especificaciones proporcionadas por SEPA.

El ajuste del modelo se examinó utilizando los criterios de información AIC (Akaike) y BIC (Schwarz), dados por las fórmulas (2) y (3), respectivamente, donde k es el número de parámetros estimados en el modelo, n el tamaño de la muestra y L corresponde al valor de la devianza (o varianza) entregadas por el programa.

$$AIC = -2 \log(L) + 2(k + 1) \quad (2)$$

$$BIC = -2 \log(L) + \ln(n)(k + 1) \quad (3)$$

Posteriormente, se estimaron las diferencias promedio entre los grupos con y sin integración según la ecuación (4):

$$\text{Logit}\{P_{ig}(\theta^*)\} = \theta^* - \delta_i + \Delta_g + \gamma_i G \quad (4)$$

Donde $P_{ig}(\theta^*)$ se refiere a la probabilidad de respuesta correcta de una persona perteneciente al grupo g en el ítem i ; θ^* representa la distribución de la capacidad de la persona examinada; δ_i es el parámetro de dificultad del ítem i ; γ_i se refiere a la presencia de DIF en el ítem i ; G indica un grupo de referencia o grupo focal que se comparan entre sí, $G = 0$ si $g = R$ (grupo de referencia), $G = 1$ si $g = F$ (grupo focal); y Δ_g es la diferencia entre la capacidad media de personas en el grupo R y F (Santelices y Wilson, 2012). Para evaluar la significación estadística de la diferencia promedio entre grupos, se prueba la hipótesis nula H_0 : no existe diferencia en la habilidad promedio de los grupos. La diferencia entre grupos distribuye como χ^2 con 1 grado de libertad. Cuando χ^2 ($p < .05$) es significativo, implica que los efectos diferenciales están presentes (Santelices y Wilson, 2012).

Para el análisis de DIF se exploró si existe diferencia en la probabilidad de respuesta correcta a un ítem i , en los grupos con y sin NEET, para un determinado grado de habilidad con el estimador de Wald y con el estadístico *t de student*. El estadístico de Wald tiene una distribución χ^2 con un grado de libertad ($p < .05$). En la prueba de Wald, la estimación de máxima verosimilitud del parámetro de interés θ , es decir $\hat{\theta}$, se compara con el valor propuesto θ_0 , bajo el supuesto de que la diferencia entre ambos será aproximadamente una distribución normal (ver ecuación 6). Típicamente, el cuadrado de la diferencia se compara con una distribución de chi-cuadrado. En el caso univariado estará dado por la ecuación (5):

$$(\hat{\theta} - \theta_0)^2 / \text{Var}(\hat{\theta}) \quad (5)$$

Alternativamente, la diferencia puede ser comparada con una distribución normal. En este caso el resultado es:

$$(\hat{\theta} - \theta_0) / \text{se}(\hat{\theta}) \quad (6)$$

Donde $\text{se}(\hat{\theta})$ es el error estándar de la estimación de máxima verosimilitud.

Dado que el estimador *t-student* involucra supuestos más flexibles que el estimador de Wald, tales como que el tamaño de muestra es grande (asumiéndose distribución normal) y que las diferencias de varianzas entre grupos son desconocidas, se evaluaron los ítems con DIF significativo de acuerdo a la ecuación 6 y $p < .05$.

Para los ítems que presentan DIF significativo se realizó un estudio del tamaño del DIF según Paek (2002) donde se clasifica como DIF imperceptible (A) si el valor estimado es menor que 0.213; presentará DIF moderado (B) si el valor está entre 0.213 y 0.319 y tendrá DIF grande o severo (C) si el valor estimado de DIF es mayor a 0.319 (Santelices y Wilson, 2012).

3.2. Análisis complementario en base a la encuesta sobre NEE

Dados los resultados inesperados obtenidos, se consideró necesario indagar cómo se diagnostica, categoriza y operacionaliza la variable Integración en los establecimientos involucrados en el estudio, con el fin de explorar algunas hipótesis que explicarían dichos resultados.

3.2.1. Participantes

Se encuestó a los coordinadores PIE y/o jefes de UTP de todos los establecimientos disponibles en la base de datos SEPA que aceptaron participar en el estudio pertenecientes a tres comunas populosas de Santiago (Maipú, Macul y Puente Alto). La ciudad de Santiago concentra la mayor parte de la población del país y estas comunas representan al 73% de los colegios de Chile (ADIMARK, 2007).

3.2.2. Instrumento

Se construyó una encuesta de 15 preguntas de respuesta cerrada que fue mejorada según las sugerencias planteadas por académicos de la universidad, un coordinador del proyecto de integración y profesionales de SEPA. Se piloteó el instrumento en dos establecimientos, uno municipal y uno subvencionado, ubicados en las mismas comunas donde se aplicó la encuesta.

3.2.3. Procedimiento

Se contactó telefónicamente a los colegios involucrados en el estudio y se coordinó la aplicación de la entrevista. En ella, cada encuestado firmó un consentimiento escrito y accedió al registro de audio.

3.2.3. Análisis de datos

El análisis de la información recogida se realizó mediante tablas de frecuencia de las respuestas, lo que permitió visualizar similitudes y diferencias.

4. Resultados

4.1. Análisis de DIF

El grado de ajuste de los datos a un modelo unidimensional y multidimensional se evaluó obteniéndose los resultados presentados en la Tabla 1, donde se compara el valor obtenido para L, AIC y BIC (ver ecuaciones 2 y 3).

Si se utiliza solo el valor de la devianza (L_{uni} versus L_{multi}), para el sector de lenguaje, los datos se ajustan mejor a un modelo unidimensional, mientras que para matemática, los datos se ajustan mejor a un modelo multidimensional. Sin embargo, al incorporar

en la evaluación el tamaño de la muestra (n) y el número de parámetros estimados en cada análisis (k), calculando AIC y BIC para todos los niveles, los datos se ajustan mejor a un modelo unidimensional, dado que un valor menor de L, AIC o BIC implica un mejor ajuste al modelo (Santelices y Wilson, 2012).

Tabla 1. Comparación de resultados ajuste a modelo unidimensional y multidimensional usando L, AIC y BIC

PRUEBA	K UNI	L UNI	AIC UNI	BIC UNI	K MULTI	N	L MULTI	AIC MULTI	BIC MULTI
4L	41	139992.5	73.7	323.9	56	2858	140415.8	103.7	443.3
5L	41	140472.1	73.7	324.5	56	2899	141193.2	103.7	444.1
6L	41	123886.5	73.8	318.6	56	2508	125737.9	103.8	436.0
7L	51	174372.3	93.5	402.5	66	2812	176619.9	123.5	521.6
8L	51	147999.9	93.7	395.5	66	2452	148844.2	123.7	512.6
4M	41	133202.0	73.8	322.3	57	2743	133020.5	105.8	448.9
5M	41	143383.6	73.7	324.7	57	2912	144581.3	105.7	452.3
6M	41	130744.8	73.8	320.3	57	2620	130587.2	105.8	446.3
7M	51	176747.8	93.5	403.0	67	2840	176379.9	125.5	530.2
8M	51	162944.4	93.6	398.3	67	2593	162474.3	125.6	524.1

Fuente: Elaboración propia con base en la información reportada por el análisis Conquest.

4.2. Diferencia de habilidad entre grupos

Según los resultados obtenidos con el programa, para las pruebas de lenguaje y matemática aplicadas en los niveles 4°, 5°, 6°, 7° y 8° no existe diferencia significativa en la habilidad promedio entre grupos, puesto que χ^2 en todos los casos es menor que 3.841.

4.2.1. Análisis de ítems que presentan DIF

Al aplicar el criterio de significancia estadística utilizando el estadístico de Wald y *t-student* para las pruebas aplicadas, se obtiene lo reportado en la Tabla 2. Según el estadístico de Wald, no existen ítems con DIF significativo en ninguna de las pruebas aplicadas. Los resultados arrojados por el estadístico *t* en las pruebas de lenguaje aplicadas en 4° y 7° no presentan DIF, mientras que en 5°, los ítems I₃₆ e I₃₇ (5%) favorecen al grupo con integración. En 6°, I₃ perjudica a los alumnos con integración mientras que I₂₇ los favorece. En 8°, 5 ítems presentan DIF (10%); I₁, I₂ e I₇ perjudican a los niños con integración mientras que I₂₉ e I₃₃ los favorecen.

Para la prueba de matemática aplicada, en 5°, 6° y 7° no se evidencia DIF; sin embargo, en 4° los ítems I₁₆, I₂₁ e I₃₅ (7,5% del total) favorecen a los niños con integración. En 8°, se encontraron 4 ítems con DIF (8%); I₄, I₈, I₂₀ perjudicarían a los niños con integración mientras que I₂₅ los favorecería.

4.2.2. Tamaño de DIF

El tamaño del efecto se estableció según lo descrito por Paek en 2002; los resultados obtenidos se muestran en la Tabla 3.

En la prueba de lenguaje de 5° Básico, los ítems con DIF presentan un efecto moderado a favor del grupo con integración. En 6°, I₂₇ presenta un efecto grande a favor del grupo integrado, mientras que I₃ tiene un efecto moderado en contra de ese grupo. En 8° I₁ e I₃₃ presentan un efecto moderado de tendencia mixta, mientras que I₂, I₇ e I₂₉ poseen un efecto grande, también de tendencia mixta.

En la prueba de 4° matemática, I₃₅ presenta un tamaño del efecto imperceptible, I₁₆ muestra un efecto moderado e I₂₁ un efecto grande, todos a favor del grupo integrado. En 8°, I₄ e I₈ tienen un efecto moderado en contra del grupo con integración mientras que I₂₀ e I₂₅ presentan un efecto severo de tendencia mixta.

Tabla 2. Análisis de DIF en prueba SEPA lenguaje y matemática

PRUEBA	Nº DEL ÍTEM	ESTADÍSTICO CALCULADO	WALD (>3,84)	T-STUDENT (>1,96)
4M	16	2.37	no	si
	21	3.27	no	si
	35	1.99	no	si
5L	36	2.08	no	si
	37	2.59	no	si
6L	3	-2.12	no	si
	27	2.34	no	si
8L	1	-1.96	no	si
	2	-2.8	no	si
	7	-2.39	no	si
	29	2.62	no	si
	33	2.20	no	si
8M	4	-2.18	no	si
	8	-2.05	no	si
	20	-2.27	no	si
	25	2.67	no	si

Fuente: Elaboración propia con base en la información reportada por el análisis Conquest.

Tabla 3. Tamaño del efecto para los ítems con DIF significativo

PRUEBA	Nº DEL ÍTEM	ESTADÍSTICO T CALCULADO	DIF ESTIMADO	EFFECTO
5L	36	2.08	0.260	B
	37	2.59	0.280	B
6L	3	-2.12	-0.253	B
	27	2.34	0.356	C
8L	1	-1.96	-0.292	B
	2	-2.80	-0.520	C
	7	-2.39	-0.356	C
	29	2.62	0.382	C
	33	2.20	0.310	B
4M	16	2.37	0.239	B
	21	3.27	0.334	C
	35	1.99	0.203	A
8M	4	-2.18	-0.291	B
	8	-2.05	-0.297	B
	20	-2.27	-0.388	C
	25	2.67	0.355	C

Fuente: Elaboración propia con base en la información reportada por el análisis Conquest.

4.2.3. Encuesta

Los datos recopilados mediante la encuesta se observan en la Tabla 4.

Tabla 4. Resultados encuesta aplicada a establecimientos que reportaron integración en SEPA

COLEGIO	1	2	3	4	5	6
Dependencia	mun	mun	mun	ps	ps	ps
PIE	si	si	si	no	no	si
tipo NEE (NEEP, NEET)	ambas	ambas	ambas	neet	ambas	ambas
Responsable asignación	pie	utp-pie	pie	psic-orient	psic-orient	pie-coor.ped
TDA	si	si	si	si	si	si
TEL	si	si	si	no	si	si
DEA	si	si	si	si	si	si
FIL	si	si	si	si	si	si
Diag. pruebas std	si	si	si	no	no	si
pruebas (SEPA)	si	si	si	no	no	si
informe psicólogo	si	si	si	si	si	si
informe neurólogo	si	si	si	si	si	si
Informe fonoaudiólogo	si	si	si	no	no	si
anamnesis	si	si	si	si	no	si
profesor	si	si	si	si	no	si
N° integrados promedio	15	15	10	5	3	10
NEET frec (TDA)	si	no	no	si	si	si
NEET frec (DEA)	si	si	no	no	no	si
NEET frec (FIL)	si	no	si	no	no	no
REPORTA > 5	no	no	no	no	no	no
Informe SEPA-MIN igual	si	si	si	si	si	si
Adecuaciones	no	no	no	no	no	no

Fuente: Elaboración propia con base en la información reportada por la encuesta.

Notas: MUN = municipal; PS = particular subvencionado; PIE = proyecto integración escolar; UTP = Unidad Técnica Pedagógica; NEEP = necesidades educativas especiales permanentes = NEET = necesidades educativas especiales transitorias; TDA = trastorno déficit atencional; TEL = trastorno específico del lenguaje; DEA = dificultades específicas del aprendizaje; FIL = coeficiente intelectual rango límite; psic= psicólogo; orient= orientador; coor.ped = coordinador pedagógico.

4.3. Análisis de datos recopilados con la encuesta

Los resultados obtenidos a través de la encuesta evidencian que:

En 5 de los 6 los colegios donde se aplicaron las encuestas existen alumnos con NEEP y NEET. Solo en uno (subvencionado) se halló que no existen alumnos con NEEP, sino solo alumnos con NEE transitorias. En todos los establecimientos encuestados existen alumnos con todas las NEET descritas en los antecedentes: TDA, DEA y CI límite (6/6) y TEL en 5/6. En los colegios municipales se comprueba la existencia de alumnos con TDA, DEA, CI límite y TEL (3/3), algo similar ocurre en la dependencia particular subvencionada, donde TDA, DEA y CI límite se encuentra en 3/3 y TEL en 2/3. La NEET más frecuente en los colegios municipales es DEA o FIL con 2/3. En cambio, los colegios particulares subvencionados indican que es TDA (3/3).

La forma de diagnosticar las NEET se relaciona con la existencia de un proyecto de integración (PIE) en el establecimiento. En todos los establecimientos municipales (3/3) y en el único establecimiento subvencionado que cuenta con PIE (1/3), el diagnóstico de NEE es realizado por el equipo de profesionales que conforman el PIE (psicopedagogo, fonoaudiólogo, psicólogo, etc.). En los colegios del sector subvencionado que no cuentan con PIE, el diagnóstico es realizado solo por un psicólogo externo y el orientador del colegio. El Ministerio de Educación de Chile define la documentación requerida para la evaluación de los alumnos con NEET en establecimientos con PIE. Esta incluye: autorización de los padres, diagnóstico con

pruebas estandarizadas, informe psicopedagógico, informe neurológico, informe de fonoaudiólogo, anamnesis, entre otros. En los colegios particulares subvencionados que no cuentan con PIE, la evaluación se basa solo en el informe de un psicólogo y de un neurólogo, incluyendo además en 1/3 colegios anamnesis y en 2/3 la recomendación de un profesor.

El total de alumnos integrados que cada establecimiento reporta oficialmente a SEPA se relaciona con el hecho de tener o no tener PIE. Se observa que los colegios con PIE solo reportan el número máximo de alumnos con NEET que permite la ley ($N = 5$), mientras que los que no tienen PIE reportan un número de alumnos menor o igual a 5. Este fenómeno ocurre de forma indiferente de si son establecimientos municipalizados o subvencionados. Es decir, el número de alumnos integrados reportados a los administradores de la prueba SEPA en ninguno de los 6 colegios supera lo permitido por la ley (5-7). Sin embargo, en todos ellos habría entre 5 y 10 alumnos integrados por curso adicionales, que no habrían sido reportados a SEPA pero si son mencionados en la encuesta. Por otro lado, en los colegios particulares subvencionados sin PIE (2/3), la cantidad de niños con NEET reportada a SEPA coincide con el número de alumnos con NEET reportados en la encuesta.

Todos los establecimientos reportaron que no realizaron adecuaciones durante la aplicación de la prueba.

5. Discusión

En esta investigación se aborda las interrogantes de si el puntaje obtenido en estas pruebas es reflejo del grado de habilidad que el alumno posee y de si está libre de sesgo, es decir, es independiente de la existencia de necesidades educativas especiales transitorias (NEET). Los análisis de DIF realizados en las pruebas de lenguaje y matemática aplicada en los niveles desde 4° a 8° muestran que los ítems con DIF encontrados en 5°, 6° y 8° (en lenguaje) y 4° y 8° (en matemática) no superan el 10%, lo que coincidiría con los resultados de Moreira (2008), quien reportó un 18% de ítems con DIF según TDAH, y con Cho et al. (2012), quienes reportaron un 21% de ítems con DIF.

Los ítems con DIF encontrados en nuestro estudio tanto para la prueba de lenguaje como de matemática son de tamaño moderado o grande de tendencia mixta, es decir, en algunos casos favorecen a los niños con NEET y en otros los perjudican. Esto implica que no existen pruebas suficientes para declarar la existencia de sesgo (favorecer a un grupo sobre el otro) en los resultados de la prueba SEPA, lo que aporta alguna evidencia respecto de la validez de los resultados de dicha prueba.

Además, los resultados sugieren que no existe diferencia en la habilidad promedio entre los grupos estudiados, lo que concuerda con los estudios reportados por Gregg (2007) y por Cho et al. (2012). Sin embargo, no coinciden con lo reportado por Kettler et al. (2005) y Mandinach et al. (2005), puesto que dichos autores sostienen que solo si se aplican adecuaciones al grupo de niños integrados no existirían diferencias entre grupos. Con el fin de comprobar si los resultados obtenidos reflejan similitudes en la habilidad promedio de los grupos o debilidades de la información analizada, se plantearon tres posibles hipótesis. Algunas de ellas fueron exploradas con la encuesta.

La primera hipótesis consideró que los colegios que reportaron integración podrían estar realizando adecuaciones durante la aplicación de la prueba, lo que explicaría un

mejor rendimiento de los niños integrados y por ello no se observarían diferencias en el puntaje promedio entre grupos.

Una segunda hipótesis exploró la posibilidad de que los establecimientos que reportaron integración estuviesen reportando solo una parte de los alumnos, quedando el grupo sin integración con alumnos que sí presentan NEET. Esto indicaría que el reporte de alumnos con NEET por sala estaría condicionado al máximo permitido por la ley (7). En este caso, la comparación estadística entre grupos con NEET y sin NEET perdería algo de validez, ya que el grupo teóricamente sin NEET sería en realidad “híbrido”.

Una tercera posibilidad consideró que los colegios que reportaron integración hubiesen logrado que las dificultades de los alumnos con NEET fuesen superadas con el trabajo de los especialistas, lo que habría resultado en una nivelación de la habilidad promedio entre ambos grupos.

A la luz de los datos reportados por la encuesta, es posible decir que en los colegios que cuentan con PIE, el número de alumnos integrados en aula supera lo descrito por la ley. Esto implica que existió un número de alumnos integrados que no fueron reportados como tales, originándose por una parte un grupo “integrado” de menor tamaño al real y un grupo “sin integración” que contenía en realidad alumnos integrados. Este hallazgo de alguna manera invalida el análisis de DIF realizado en los colegios que, habiendo declarado alumnos integrados a SEPA, presentan también casos de niños con NEET no declarados. Consiguientemente, no es posible inferir la existencia de sesgo en la prueba ya que el comportamiento diferencial de los ítems depende fuertemente de las características de los grupos comparados.

De acuerdo con los resultados de la encuesta, se desestimó que los colegios estuvieran realizando adecuaciones al aplicar la prueba, por lo que la similitud en la habilidad promedio observada no sería debido a esta causa.

Por medio de la encuesta no solo se indagó información relativa a las hipótesis anteriormente mencionadas, sino que se exploraron otros aspectos entre los que podemos mencionar los siguientes:

En primer lugar, los datos recopilados a través de la encuesta permiten corroborar la existencia de NEET y NEEP en las aulas, comprobando que la integración es un hecho patente en los establecimientos del país. En segundo lugar, se evaluó lo descrito por el Mineduc respecto del porcentaje de cada NEET según dependencia en el país. Para el sector municipal, se evidencia que las NEET más frecuentes son DEA y FIL, lo que coincide con los datos disponibles, mientras que en el sector particular subvencionado existen tanto TDA como DEA, pero solo uno de los colegios subvencionados encuestados señala que en su comunidad la NEET más frecuente es DEA, lo que coincidió parcialmente con los datos disponibles, ya que los datos reflejan que la NEET más frecuente es TDA (3/3).

Los datos recolectados respecto de la documentación requerida para diagnosticar a un alumno con NEET y del protocolo de diagnóstico para dichos alumnos descrita por los establecimientos municipales y subvencionados que cuentan con PIE encuestados coinciden íntegramente con lo establecido por el Mineduc. Sin embargo, en los colegios particulares subvencionados que no han implementado PIE existe mayor libertad en cuanto a quién realiza la asignación y con base en qué tipo de documentación. En ellos, se constató que solo se requiere de informe psicopedagógico y/o neurológico.

Cabe destacar que los resultados reportados por el análisis complementario tienen una generalizabilidad limitada debido principalmente al tamaño de la muestra y a que la voluntariedad en la participación puede haber introducido sesgos en las respuestas. Sin embargo, es importante considerar que ella incluye representantes de ambos tipos de dependencia administrativa (municipal y subvencionada) y considera los niveles socioeconómicos más frecuentemente observados en las escuelas de Chile (73% del país), lo que entrega valor a los resultados.

6. Conclusión

A partir del análisis estadístico, se observa que los ítems que presentan DIF en las pruebas de matemática y lenguaje aplicadas no superan el 10% y tienen una tendencia mixta, beneficiando a veces al grupo focal y otras veces al de referencia. Los resultados obtenidos evidencian, sin embargo, que no existiría diferencia en la habilidad promedio entre los grupos estudiados. Dicho resultado coincide solo parcialmente con los resultados de investigaciones anteriores, por lo que se exploró, mediante una encuesta, el posible uso de adecuaciones y la forma en que las escuelas estaban reportando las NEE de sus alumnos.

A partir de los datos recogidos con la encuesta fue posible constatar que los establecimientos no habían realizado ningún tipo de adecuación durante la aplicación de la prueba, por lo que la similitud estadística en la habilidad promedio estimada para el grupo con NEET y sin NEET no sería producto de ello. Aun más, los datos sustentan la hipótesis de que el grupo de referencia utilizado en el análisis DIF sería un grupo de referencia (sin integración) híbrido, ya que los colegios que presentan PIE señalan que el número de alumnos reportados a SEPA coincide con el máximo permitido por ley y que habrían existido alumnos integrados no reportados como tales. A la luz de lo anterior, no es posible inferir respecto de las posibles diferencias en la habilidad promedio entre los grupos de la prueba SEPA según NEET aplicada en lenguaje y matemática en 2012 para los establecimientos municipales y subvencionados con PIE que reportaron menor número de alumnos que los que existían en realidad en aula. Adicionalmente, aunque los ítems con DIF no superan el 10%, no existe claridad en la verdadera causa de estos hallazgos, pudiendo ser por diferencias estadísticas en el comportamiento del ítem o por deficiencias en la asignación de grupos.

Una futura investigación debería considerar un análisis de la existencia de DIF según NEET asociado a encuestas y trabajo de campo que permita asegurar que el grupo sin integración no es híbrido, es decir, que lo declarado por los establecimientos en la prueba aplicada coincide íntegramente con lo evidenciado en aula y no solo es coincidente con lo que se espera desde lo legal.

Referencias

- Adams, R., Wilson, M. y Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23. doi:10.1177/0146621697211001.
- ADIMARK. (2007). *Mapa socioeconómico de Chile: Nivel socioeconómico de los hogares del país basado en datos del Censo*. Recuperado de http://www.adimark.cl/medios/estudios/Mapa_Socioeconomico_de_Chile.pdf
- Baena, G. (1998). *Técnicas de investigación*. Ciudad de México: Editores Mexicanos Unidos.

- Bond, T. y Fox, C. (2001). *Applying the Rasch model*. Mahwah, NJ: Lawrence Erlbaum.
- Cahalan-Laitusis, C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education*, 23(2), 153-167. doi:10.1080/08957341003673815.
- Cawthon, S. (2010). Examining student factors in sources of setting accommodation DIF. *Educational and Psychological Measurement*, 74(5), 759-794. doi:10.1107/0013164413514053.
- Cho, H., Lee, J. y Kingston, N. (2012). Examining the effectiveness of test accommodation using DIF and a mixture IRT model. *Applied Measurement in Education*, 25(4), 281-304. doi:10.1080/08957347.2012.714682.
- Cohen, A., Gregg, N. y Deng, M. (2005). The role of extended time and item content on a High-Stakes Mathematics Test. *Learning Disabilities Research & Practice*, 20(4), 225-233. doi:10.1111/j.1540-5826.2005.00138.x.
- CONEVAL. (2012). *Informe de pobreza y evaluación en el estado de Sonora*. Recuperado de <http://desarrollosocial.guanajuato.gob.mx/coneval/informe-sonora.pdf>
- CONEVAL. (2012). *Porcentaje de población en situación de pobreza extrema según entidad federativa*. Recuperado de http://www.coneval.gob.mx/SalaPrensa/Documents/Comunicado005_Medicion_pobreza_2014.pdf
- Creswell, J. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3ª ed.). California, CA: Pearson.
- Delval, J. (1992). *Aprender a aprender. El desarrollo de la capacidad de pensar*. Madrid: Alhambra Longman.
- Elosúa, P. (2003). Sobre la validez de los test. *Psicothema*, 15(2), 315-321.
- Finch, W. y Finch, M. (2013). Investigation of specific learning disability and testing accommodations based differential item functioning using a multilevel multidimensional mixture item response theory model. *Educational and Psychological Measurement*, 73(6), 973-993.
- Gómez-Benito, J., Hidalgo, M. y Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles del Psicólogo*, 31(1), 75-84.
- Gregg, N. (2007). Underserved and unprepared: Postsecondary learning disabilities. *Learning Disabilities Research & Practice*, 22(4), 219-228. doi:10.1111/j.1540-5826.2007.00250.x.
- Hen, M. y Goroshit, M. (2014). Academic procrastination, emotional intelligence, academic self-efficacy, and GPA: A comparison between students with and without learning disabilities. *Journal of Learning Disabilities*, 47(2), 116-124. doi:10.1177/0022219412439325.
- Idan, O. y Margalit, M. (2014). Socioemotional self-perceptions, family climate, and hopeful thinking among students with learning disabilities and typically achieving students from the same classes. *Journal of Learning Disabilities*, 47(2), 136-152. doi:10.1177/0022219412439608.
- INEGI. (2014, 27 de octubre). *Consulta interactiva de datos*. Recuperado de <http://www.inegi.org.mx/sistemas/olap/proyectos/bd/consulta.asp?p=17118yc=27769ys=est>
- Infante, M., Matus, C. y Vizcarra, R. (2010). Desafíos a la formación docente: Inclusión educativa. *Estudios Pedagógicos*, 36(1), 287-297. doi:10.4067/s0718-07052010000100016.

- Kettler, R. J., Niebling, B. C, Mroch, A. A., Feldman, E. S., Newell, M. L., Kratochwill, T. R. ... y Bolt, D. M. (2005). Effects of testing accommodations on Math and Reading scores: An experimental analysis of the performance of students with and without disabilities. *Assessment for Effective Intervention*, 31(1), 37-48. doi:10.1177/073724770503100104.
- Lin, P. y Lin, Y. (2014). Examining student factors in sources of setting accommodation DIF. *Educational and Psychological Measurement* 74(5), 759-794. doi:10.1177/0013164413514053.
- Mandinach, E. B., Bridgeman, B., Cahalan-Laitusis, C. y Trapani, C. (2005). The impact of extended time on SAT test performance. *The College Board*, 2, 1-35.
- Meltzer, Y. (2006). תמיכה מרכזי-תל האקדמית במכללה למידה לקויות עם לסטודנטים . [A support center for learning disabled students in Tel-Hai Academic College]. *Issues in Special Education & Rehabilitation*, 21, 98-101.
- Ministerio de Educación de Chile. (2009). *Decreto N° 170/09. Normativa Educación Especial*. Recuperado de <http://www.mineduc.cl/usuarios/edu.especial/doc/201304231500550.DEC200900170.pdf>
- Moreira, T. E. (2008). Funcionamiento diferencial del ítem en pruebas de matemática para educación media. *Actualidades en Psicología*, 22(109), 91-113. doi:10.15517/ap.v22i109.16.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Muñiz, J. (2010). Las teorías de los test: Teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Navas, M. J. (2012). La medición en el ámbito educativo. *Psicología Educativa*, 18(1), 15-28. doi:10.5093/ed2012a2.
- Organización de las Naciones Unidas. (2008). *Declaración Universal de los Derechos Humanos*. Recuperada de en <http://www.temoa.info/es/node/19618>
- Paek, I. (2002). *Investigation of differential item functioning: Comparisons among approaches, and extension to a multidimensional context*. (Tesis doctoral no publicada). Universidad de California.
- Pontificia Universidad Católica de Chile. (2014, 23 de septiembre). *Sistema de Evaluación de Progreso del Aprendizaje: Pruebas SEPA*. Recuperado de <http://www.sepauc.cl/wp-content/uploads/2014/05/20140923-Sistema-Evaluacion.pdf>
- Prieto, G. y Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Ramos, R. Y. (2001). *Educación integral. Una educación holística para el siglo XXI* (Vol I). Madrid: Desclée de Brouwer.
- Santelices, M. V. y Wilson, M. (2012). On the relationship between differential item functioning and item difficulty: An issue of methods? Item response theory approach to differential item functioning. *Educational and Psychological Measurement*, 72(1), 5-36. doi:10.1177/0013164411412943.
- Sireci, S., Scarpati, S. y Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490. doi:10.3102/00346543075004457.
- Tenorio, S. (2011). Formación inicial docente y necesidades educativas especiales. *Estudios Pedagógicos*, 37(2), 249-265. doi:10.4067/s0718-07052011000200015.
- Wu, M., Adams, R. J. y Wilson, M. (1998). *ACER-ConQuest*. Hawthorn, Victoria: ACER Press.