



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

APORTES DEL ANÁLISIS PSICOMÉTRICO A LA COMPRENSIÓN DE LA
ESTRUCTURA DEL CONOCIMIENTO MATEMÁTICO PARA ENSEÑAR

TESIS PARA OPTAR AL TÍTULO DE INGENIERA CIVIL MATEMÁTICA

VALENTINA SOFÍA GIACONI SMOJE

PROFESORA GUÍA:
MARÍA LEONOR VARAS SCHEUCH

MIEMBROS DE LA COMISIÓN:
NANCY LACOURLY VENTRE
SALOMÉ MARTÍNEZ SALAZAR

Este trabajo ha sido parcialmente financiado por FONDECYT 1090292

SANTIAGO DE CHILE
DICIEMBRE 2012

Resumen

A nivel internacional se ha caracterizado una estructura del conocimiento matemático de los profesores que es específico de la tarea de enseñar matemática y se ha acumulado evidencia en cuanto a que explica significativas ganancias de aprendizaje en los alumnos. En esta estructura se distinguen seis factores que se agrupan en dos categorías: el *conocimiento disciplinar* y el *conocimiento pedagógico del contenido*. En esta memoria se trabajó con los datos resultantes de la aplicación de un instrumento diseñado para medir el *conocimiento de alumnos y matemáticas* (CAM), que forma parte del *conocimiento pedagógico del contenido*. Además se utilizaron los datos resultantes de la aplicación simultánea de una prueba internacional que evalúa dos componentes del *conocimiento disciplinar*: el *conocimiento matemático común* y el *conocimiento matemático especializado*, para compararlos con los de la prueba CAM, y aportar así tanto a su validez como a la comprensión de la estructura del conocimiento mencionado.

Por otra parte para determinar la validez y confiabilidad de las mediciones de un constructo teórico es necesario entender, utilizar y analizar los resultados de diversos modelos matemáticos. En la presente memoria primero se estudiaron y describieron estos modelos y luego se aplicaron a las dos pruebas mencionadas.

En la introducción de esta memoria se describe el marco teórico del *conocimiento matemático* de los profesores que es específico de la tarea de enseñar . En el segundo capítulo se describen los métodos de análisis factoriales exploratorio y confirmatorio. En la parte de resultados de este capítulo se muestra la aplicación de los métodos de análisis factorial exploratorio para determinar las dimensiones que mide la prueba CAM y realizar una selección de ítems que sea unidimensional. También se utilizó el método de análisis factorial confirmatorio para confirmar la hipótesis de que los constructos *conocimiento matemático común y específico* y *conocimiento de alumnos y matemáticas* son distinguibles. En el tercer capítulo se describe la teoría clásica de test con especial énfasis en el concepto de confiabilidad y su estimación. En la sección de resultados se determina la confiabilidad de la prueba CAM total y la selección de ítems. Se presenta también el cálculo y análisis de los estadísticos de los ítems. En el cuarto capítulo se describe la Teoría de Respuesta al Ítem (TRI o IRT). Primero se describen dos modelos de la TRI: el modelo de un parámetro (Rasch) y el de dos parámetros. Se presenta la aplicación de estos modelos a la prueba CAM, en particular se describe la relación entre los ítems y las personas.

Por último se presentan las conclusiones. Se obtuvo una prueba que permite evaluar el *conocimiento de alumnos y matemáticas* con un buen nivel de confiabilidad y validez.

Agradecimientos

A mi Máma, Pápa y Hermana que me han apoyado siempre en todos los ámbitos de la vida, sin ustedes esto no existiría. Gracias por todo su amor, compañía y apoyo. También les agradezco todo lo que me han enseñado, especialmente saber las cosas que importan y que hay que esforzarse para lograr las metas.

A la profesora Ximena, que me enseñó matemáticas en el colegio y me dejó la gran interrogante de que era lo que necesitaba un profesor de matemáticas para enseñar tan bien como ella.

A la profesora Salomé Martínez por lo que me ha enseñado y porque es un gran ejemplo y por su preocupación por el ambiente del DIM.

A la profesora Nancy Lacourly por sus generosidad para entregar todo lo que sabe, por su sencillez y simpatía. También agradezco todos los conocimientos estadísticos que me ha entregado en todos los ramos que me ha enseñado y en particular por su tiempo y dedicación a esta memoria.

A la profesora Leonor Varas por haber confiado en mí todos estos años y haberme dado muchísimas posibilidades de aprender y crecer. Por permitirme entrar al mundo de la investigación en educación. Por ser tan acogedora, escucharme y enseñarme muchas cosas de matemática, educación, y de la vida.

A Eterin, Sra Silvia, Sra Gladys, Don Oscar y Don Luis Mella que siempre nos han apoyado y acompañado.

A todas las personas que forman al preuniversitario José Carrasco Tapia por darme la oportunidad de aprender a enseñar y conocer a mucha gente maravillosa.

A mis amigos de la U que conozco del primer año a la Cata, Nicol, Pato, Juanano, Caracola y Gato. A mis amigos de matemática, sin ustedes no hubiera llegado hasta aquí, en especial a la Sandris, Poncho, Pedro, Tata, Monique, Abelin, Johan, Poly, Riffo, Felipe, Zuni, Pasti, Conti, Chiri, Pablo, Mauro. A mis amigas de la vida Sarí, Trino, Cata y Pauli que siempre me han apoyado, apañado y escuchado.

A Emilio mi amor por corregir mi memoria. Por escucharme, desestresarme, acompañarme, apoyarme y hacerme muy feliz. También le agradezco su generosidad, sabiduría, comprensión, amor y paciencia. Por último le agradezco todos sus consejos y su buena voluntad.

Índice general

Índice general	iii
Índice de tablas	v
Índice de figuras	vii
1. Introducción	1
1.1. Objetivos	2
1.2. Antecedentes Previos	3
1.2.1. Conocimiento Matemático para Enseñar	3
1.2.2. Muestra	4
2. Análisis factoriales exploratorio y confirmatorio	6
2.1. Análisis Factorial	7
2.1.1. Descomposición de la varianza de las variables observadas	8
2.1.2. Relación entre las correlaciones de las variables y los parámetros del modelo	9
2.1.3. Identificación del modelo	10
2.1.4. Estimación de las comunalidades y la matriz de cargas	10
2.1.5. Elección del método	14
2.1.6. Rotaciones	14
2.1.7. Elección del número de factores o componentes	16
2.2. Análisis de Componentes Principales	18
2.2.1. Componentes principales como combinaciones lineales de los datos que tienen máxima varianza	18
2.2.2. Relación entre el análisis factorial usual y el análisis de componentes principales	21
2.3. Análisis Factorial Confirmatorio	23
2.3.1. Notación utilizando grafos	24
2.3.2. Índices de ajuste del modelo	25
2.4. Caso de datos dicotómicos	27
2.4.1. Correlaciones Tetracóricas	27
2.5. Resultados sobre las pruebas CAM y LMT	29
2.5.1. Análisis factorial de la prueba CAM	29
2.5.2. Análisis factorial confirmatorio sobre la prueba CAM y la prueba LMT	38
3. Teoría Clásica de Test	42

3.1.	Teoría Clásica en test formados por ítems	42
3.1.1.	Descripción del modelo	42
3.1.2.	Confiabilidad en pruebas formadas por ítems	45
3.1.3.	Errores	50
3.1.4.	Estimación de la confiabilidad	50
3.1.5.	Testeo de las hipótesis de teoría clásica con análisis factorial confirmatorio	51
3.1.6.	Estadísticos de los ítems	52
3.2.	Teoría clásica de test aplicada a la prueba CAM	54
3.2.1.	Estadísticos descriptivos	54
3.2.2.	Confiabilidad	54
3.2.3.	Estadísticos de los ítems: Dificultad y Discriminación	57
4.	Teoría de Respuesta al Ítem	60
4.1.	Supuestos de los modelos	63
4.2.	El modelo de Rasch	65
4.2.1.	Mapas ítems-personas	65
4.2.2.	Ajuste de las dificultades y habilidades en el modelo de Rasch	66
4.3.	El modelo de dos parámetros	68
4.3.1.	Relación con análisis factorial	68
4.4.	Estimación de los parámetros	72
4.4.1.	Estimación de habilidad suponiendo los parámetros de los ítems conocidos	72
4.4.2.	Estimación de los parámetros de los ítems y las personas simultáneamente	73
4.5.	Función de información y error estándar de medición	74
4.6.	Aplicación de la teoría de respuesta al ítem a la prueba CAM	76
4.6.1.	Estimación del modelo de Rasch	77
4.6.2.	Estimación del modelo de 2PL	80
4.6.3.	Análisis de la función de información	81
5.	Conclusiones	84
	Bibliografía	87
A.	Resultados análisis factorial	91
A.1.	Resultados del análisis de factores principales sobre la matriz de correlaciones de Pearson	91
A.2.	Resultados del ACP sobre la matriz de correlaciones tetracóricas	94
A.3.	Resultados del análisis análisis factorial estimado con máxima verosimilitud sobre la matriz de correlaciones de Pearson	97
A.4.	Resultados del análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas	102

Índice de tablas

1.1.	Distribución de profesores y estudiantes en la muestra	5
1.2.	Distribución de las escuelas en la muestra y el universo según la dependencia administrativa y nivel de logro SIMCE	5
2.1.	Correlaciones entre las columnas de las matrices de estructura entregadas por los métodos ACP, MV, EP y MinRes aplicados a la matriz de correlaciones de Pearson.	31
2.2.	Correlaciones entre las columnas de las matrices de estructura entregadas por los métodos ACP, MV, EP y MinRes aplicados a la matriz de correlaciones tetracóricas.	32
2.3.	Para cada modelo factorial estimado se tomaron los 30 ítems que correlacionaron más alto con el factor 1. En la tabla se presenta el mínimo de estas correlaciones en cada modelo (los modelos se identifican con el método y la matriz usada en la estimación).	35
2.4.	Correlación de cada ítem con el factor 1 calculado con un ACP sobre la matriz de correlaciones de Pearson.	37
2.5.	Correlación de cada ítem con el factor 1 calculado con un ACP sobre la matriz de correlaciones tetracóricas.	37
2.6.	Promedio de los índices de ajuste en las 100 generaciones aleatorias de parcelas para cada modelo.	39
3.1.	Estadísticos descriptivos del puntaje de la prueba CAM total (61 ítems) y de la prueba CAM depurada (28 ítems)	54
3.2.	Estadísticos de ajuste en los modelos para testear las hipótesis de ítems paralelos, tau-genéricos y congenéricos en la prueba CAM total (61 ítems)	55
3.3.	Estadísticos de ajuste en los modelos para testear las hipótesis de ítems paralelos, tau-genéricos y congenéricos en la prueba CAM depurada (28 ítems)	56
3.4.	Coeficientes de confiabilidad y errores estándares de medición en ambas pruebas	56
3.5.	Dificultades de los ítems de la prueba CAM total	57
3.6.	Dificultades de los ítems de la prueba CAM depurada	57
3.7.	Discriminaciones de los ítems de la prueba CAM total	59
3.8.	Discriminaciones de los ítems de la prueba CAM depurada	59
4.1.	Rangos de ajuste para los estadísticos Infit y Outfit, tomados de [7].	67
4.2.	Niveles referenciales de discriminación dados por Baker en [2]	69
4.3.	Funciones de información de ítem y de test en el modelo de Rasch y de 2PL	74
4.4.	Estadísticos de los ítems: dificultad, Infit y Outfit para la prueba CAM total	77

4.5.	Estadísticos de los ítems: dificultad, Infit y Outfit para la prueba CAM depurada	79
4.6.	Dificultades estimadas en el modelo de dos parámetros para la prueba CAM total	80
4.7.	Discriminaciones estimadas en el modelo de dos parámetro para la prueba CAM total	81
4.8.	Dificultades estimadas en el modelo de dos parámetros para la prueba CAM depurada	81
4.9.	Discriminaciones estimadas en el modelo de dos parámetros para la prueba CAM depurada	82
4.10.	Estadísticos descriptivos de la habilidad y la información en el modelo de Rasch para la prueba CAM total y depurada	82
4.11.	Estadísticos descriptivos de la habilidad y la información en el modelo de dos parámetros para la prueba CAM total y depurada	82
A.1.	Valores propios del ACP sobre la matriz de correlaciones de Pearson	91
A.2.	Columnas ordenadas de la matriz de estructura producida por el ACP sobre la matriz de correlaciones de Pearson	92
A.3.	Valores propios del ACP sobre la matriz de correlaciones tetracóricas	94
A.4.	Columnas ordenadas de la matriz de estructura producida por el ACP sobre la matriz de correlaciones tetracóricas	95
A.5.	Valores propios del análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson	97
A.6.	Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson	98
A.7.	Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson después de una rotación promax	100
A.8.	Valores propios del análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas	102
A.9.	Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas	103
A.10.	Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas después de una rotación promax	105

Índice de figuras

1.1. Modelo sobre el <i>Conocimiento Matemático para Enseñar</i> desarrollado por el proyecto LMT (Aprendiendo Matemáticas para Enseñar) de la Universidad de Michigan [3].	3
2.1. Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones de Pearson.	30
2.2. Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones tetracóricas.	31
2.3. Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones de Pearson con la selección de ítems.	36
2.4. Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones tetracóricas con la selección de ítems.	36
2.5. Proyección de los ítems de la prueba LMT sobre el ACP de la prueba CAM.	38
2.6. Modelos de análisis factorial confirmatorio.	40
2.7. Distribución de los índices de ajuste en el Modelo 1 y el Modelo 2, para las 100 parcelas aleatorias que se estimaron.	41
3.1. Modelos de análisis factorial confirmatorio para testear las hipótesis: Los ítems son paralelos; los ítems son tau-equivalentes o los ítems son congenéricos	52
3.2. Histograma de los puntajes de la prueba CAM total y de la prueba CAM depurada.	54
3.3. Histograma de las dificultades de los ítems de la prueba CAM total y de la prueba CAM depurada	58
4.1. Curvas características del ítem para el ítem más fácil (a) y difícil (b) de la prueba CAM depurada.	65
4.2. Curvas características del ítem para dos ítems con dificultades similares, donde el primero es muy discriminante y el segundo muy poco.	68
4.3. Gráfico ítems-personas para la prueba CAM total.	78
4.4. Gráfico ítems-personas para la prueba CAM depurada.	79
4.5. Funciones de información e histogramas de la habilidad para la prueba CAM total y depurada en el modelo de Rasch.	83
4.6. Funciones de información e histogramas de la habilidad para la prueba CAM total y depurada en el modelo de dos parámetros.	83

Capítulo 1

Introducción

El éxito en el aprendizaje de la matemática a nivel escolar es un gran desafío que pocos países han podido lograr de forma satisfactoria. Parte importante en el desarrollo de este tema es la preparación y conocimientos que deben tener los profesores. Aunque es evidente que los profesores tienen que manejar los contenidos que enseñan, ha habido mucho trabajo teórico para delinear los contenidos que tiene que dominar un profesor para desempeñarse de forma adecuada. Una contribución decisiva en este tema la hizo Shulman en 1986 [36] cuando introdujo el concepto de *conocimiento pedagógico del contenido (CPC)*. Este término se refiere a aquellos conocimientos pedagógicos que son necesarios para enseñar contenidos disciplinarios. Este concepto fue un aporte mayor ya que conectó conocimiento de los contenidos y la práctica de enseñarlos, y ha generado un impacto tremendo en investigadores en educación de diversas disciplinas, que han seguido desarrollando el trabajo iniciado por Shulman.

Uno de los grupos de investigadores que han trabajado a partir de lo desarrollado por Shulman, pertenece a la Universidad de Michigan, y se agrupa en torno al proyecto Learning Mathematics for Teaching (LMT). Ellos desarrollaron un marco conceptual y definieron el *conocimiento matemático para enseñar (CME)* como el conocimiento matemático que se necesita para la tarea de enseñar, el cual está formado por una dimensión puramente disciplinar y otra dimensión de tipo pedagógica. Este equipo y otros, han demostrado que la presencia de este contenido genera logros de aprendizaje significativos en los estudiantes, más que ninguna de muchas otras variables [20] y [5]. Dentro de las componentes del *conocimiento matemático para enseñar* distinguieron el *conocimiento de alumnos y matemáticas* que se refiere al conocimiento de los alumnos como aprendices de matemática.

En Chile, dentro del proyecto Fondecyt 1090292 que auspicia a esta memoria se desarrolló un instrumento nuevo que consiste en una prueba con ítems de alternativas para medir el *conocimiento de alumnos y matemáticas*. Este conocimiento no ha podido ser medido de manera exitosa en trabajos anteriores del grupo de LMT [22]. El instrumento, que llamamos prueba CAM, es el objeto de análisis principal de esta memoria, cuyo propósito es estudiar y aplicar métodos estadísticos de manera crítica para analizar sus características técnicas y validarlo.

Además de la prueba CAM, se utilizó una prueba diseñada por el equipo LMT, que evalúa otra componente del CME llamada *conocimiento matemático común y específico*. A esta prueba la llamamos prueba LMT. La utilización de ambas pruebas se hizo para poder relacionar sus mediciones. Ambas pruebas se aplicaron a una muestra de profesores de educación básica y estudiantes de pedagogía básica.

Para poder evaluar constructos que son imposibles de medir directamente como un conocimiento o característica psicológica, es necesario verificar que el instrumento es confiable, es decir que mide con precisión y que tiene validez. Que un instrumento tenga validez significa que hay evidencia de que mide el constructo teórico. La prueba de LMT ya está validada y por eso podemos usar sus mediciones para relacionar al *conocimiento matemático común y específico* con lo que mide la prueba CAM. Sin embargo, ésta última por ser una prueba nueva tiene que ser analizada con diversos métodos para asegurar su confiabilidad y validez.

Lo que se hizo en esta memoria fue aplicar a la prueba CAM tres métodos estadísticos: Análisis Factorial, Teoría Clásica de Test y Teoría de Respuesta al Ítem para:

1. Identificar el constructo que está midiendo el instrumento y verificar que coincide con el constructo teórico *conocimiento de alumnos y matemáticas*.
2. Analizar la relación del *conocimiento matemático común y específico* con el *conocimiento de alumnos y matemática*.
3. Verificar que la prueba tenga confiabilidad, es decir, que tenga buena calidad como instrumento de medición.
4. Evaluar la discriminación y distribución de dificultades de los ítems de la prueba CAM.

El cumplimiento de estos puntos permite utilizar el instrumento para evaluar de manera confiable la presencia del constructo en la población de profesores y alumnos de pedagogía en enseñanza básica. Además un análisis detallado del instrumento consigue que esta evaluación tenga una alta riqueza descriptiva. También el medir de manera confiable el constructo permite relacionar su medición con otras variables de interés.

1.1. Objetivos

El objetivo principal de esta memoria es analizar la prueba CAM para entregar evidencia respecto a la dimensión que mide y a su confiabilidad. Para hacer esto se van a utilizar tres teorías o métodos estadísticos: Análisis Factorial, Teoría Clásica de Test y Teoría de Respuesta al Ítem. Los objetivos específicos son:

1. Estudiar y describir los métodos utilizados.
2. Utilizar los tres métodos para el análisis y evaluación de la prueba CAM.
3. Establecer las relaciones teóricas entre los métodos y analizar cómo se reproducen en las estimaciones y resultados obtenidos a partir de la prueba CAM.

1.2. Antecedentes Previos

1.2.1. Conocimiento Matemático para Enseñar

El marco conceptual del conocimiento matemático para enseñar se resume en el esquema de la Figura 1.1.

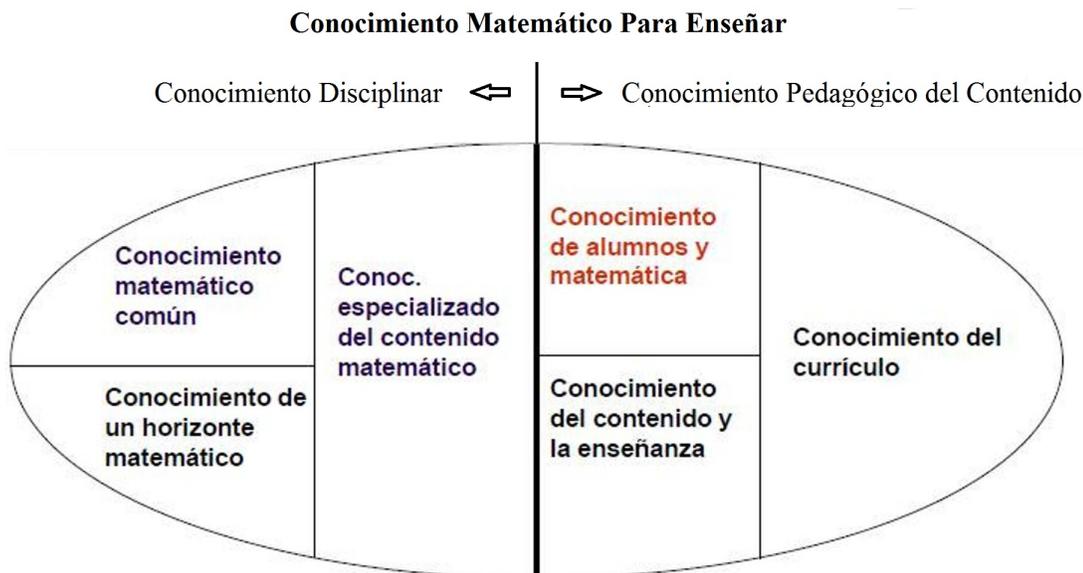


Figura 1.1: Modelo sobre el *Conocimiento Matemático para Enseñar* desarrollado por el proyecto LMT (Aprendiendo Matemáticas para Enseñar) de la Universidad de Michigan [3].

En este esquema se distinguen dos categorías principales que forman al CME, el *conocimiento disciplinar* y el *conocimiento pedagógico del contenido*. La característica principal de estos dos conocimientos es que están contextualizados a la tarea de enseñar y a la matemática que se enseña. Esto significa que el *conocimiento disciplinar* se refiere a conocimientos de matemática, pero que están elegidos en relación a lo que se enseña y a como enseñarlo. Por otro lado el *conocimiento pedagógico del contenido* es un conocimiento pedagógico sobre los estudiantes, la enseñanza y el currículo conectado con la matemática que se va a enseñar. No se refiere a un conocimiento pedagógico general o aparte del contenido.

A continuación se describe cada una de las seis componentes que forman al *conocimiento matemático para enseñar*:

1. *Conocimiento matemático común*: Este conocimiento se refiere a manejar contenido matemático que es común en muchas disciplinas, es decir que no es específico de la tarea de enseñar. Se trata principalmente de que los profesores puedan resolver los problemas que asignan a sus alumnos. Un ejemplo de este conocimiento es usar correctamente algoritmos para restar, multiplicar, dividir, etc. Este conocimiento es esencial para la enseñanza [3].
2. *Conocimiento especializado del contenido matemático*: Este conocimiento se refiere a la matemática que tiene que saber un profesor para enseñar y que es específica del profesor

(por ejemplo un matemático o un contador no tendrían porqué conocerla). Un ejemplo de este conocimiento es determinar si un algoritmo inventado por un niño sirve siempre o sólo en el ejemplo usado. Otro ejemplo es el conocer los “porqué” de las propiedades y los algoritmos [3].

3. *Conocimiento de un horizonte matemático*: Este conocimiento se refiere a saber que temas matemáticos vienen después de un curso, para poder sentar bases adecuadas para que el aprendizaje se logre. También se refiere a saber que conexiones o relaciones entre las materias van a servir en años siguientes [3].
4. *Conocimiento de alumnos y matemática*: Este conocimiento mezcla saber sobre los estudiantes y sobre la matemática. Por ejemplo se refiere a saber si un ejemplo escogido va a motivar o no a los estudiantes, anticipar que va a ser fácil o difícil al resolver una tarea. Otros ejemplos de la presencia de este conocimiento es la capacidad de interpretar pensamientos incompletos o imprecisos expresados por los estudiantes o saber la concepciones correctas y erradas que suelen tener ellos. También el conocer errores frecuentes de los estudiantes y como se asocian a determinadas estrategias de instrucción. Este conocimiento se usa en tareas que necesitan atención al contenido y a aspectos particulares de los estudiantes [22].
5. *Conocimiento del contenido y la enseñanza*: Este conocimiento combina el saber sobre enseñar y el saber matemáticas. Muchas de las tareas de enseñar consisten en un conocimiento matemático y del diseño de la instrucción. Por ejemplo como secuenciar un contenido, que ejemplos elegir, que modelos o metáforas usar etc. [3].
6. *Conocimiento del currículo*: Esta parte del conocimiento pedagógico del contenido se refiere a conocer el material curricular dado para enseñar un determinado tema en un determinado nivel. También se refiere a conocer los materiales instruccionales asociados a una determinada parte del currículo. Por último se refiere también a conocer los efectos que podría tener el uso de una parte específica del currículo o de los materiales para circunstancias determinadas [3].

Aunque se ha encontrado evidencia de que el conocimiento matemático para enseñar es multidimensional, no es el objetivo encontrar dimensiones que coincidan precisamente con esta conceptualización teórica [3]. Sin embargo debieran de existir dimensiones posibles de distinguir. En particular en esta memoria se va a testear la hipótesis de que, a pesar de estar relacionados, el *conocimiento matemático común y específico* se puede distinguir del *conocimiento de alumnos y matemática*. El probar esta hipótesis tiene importancia para posteriormente desarrollar políticas para la formación de profesores [21].

1.2.2. Muestra

Las pruebas CAM y LMT se aplicaron durante los años 2010 y 2011 en las ciudades de Santiago, Concepción, San Felipe y Lolol a personas que no participaron en las etapas previas de generación y validación de ítems.

Los datos analizados de las pruebas CAM y LMT, pertenecen a la población objetivo del estudio, que son los profesores y estudiantes de enseñanza básica. Entre los profesores y estudiantes se distinguen a los que hicieron o están haciendo una mención en matemáticas

de los que no. La distribución de profesores y estudiantes con y sin mención es la siguiente:

	N
Profesores de enseñanza general básica	148
Profesores de enseñanza básica, especialistas en matemáticas	61
Estudiantes de pedagogía en enseñanza básica	66
Estudiantes de pedagogía en enseñanza básica con especialidad en matemáticas	12
TOTAL	287

Tabla 1.1: Distribución de profesores y estudiantes en la muestra

Los profesores de la muestra provienen de diversos colegios, en el aspecto socioeconómico y en rendimiento según el nivel de logro del SIMCE en matemáticas (considerando los datos para 4° básico). La distribución de los colegios según estos indicadores se muestra en la Tabla 1.2. Dentro de cada casillero de la tabla se encuentra el número de establecimientos de la muestra y entre paréntesis el tamaño del universo correspondiente. Los estudiantes que respondieron la prueba pertenecían a la Universidad de Concepción.

Dependencia administrativa del establecimiento	SIMCE Inicial	SIMCE In-termedio	SIMCE Avanzado	TOTAL
Municipal	9 (591)	14 (516)	0 (19)	23(1126)
Particular Subvencionado	2 (351)	22 (925)	5 (147)	29(1423)
Particular Pagado	0 (7)	1 (81)	5 (189)	6(277)
TOTAL	11 (949)	37(1522)	10(355)	58 (2826)

Tabla 1.2: Distribución de las escuelas en la muestra y el universo según la dependencia administrativa y nivel de logro SIMCE

Capítulo 2

Análisis factoriales exploratorio y confirmatorio

El análisis factorial exploratorio se utiliza para encontrar factores que expliquen a un conjunto de variables. En general el número de factores necesario para explicar las variables observadas suele ser mucho menor al total de variables y permite determinar cuales son los elementos que están generando variabilidad en los datos. Por otra parte, el análisis factorial confirmatorio permite proponer una estructura de factores que explican los datos y testear si esta estructura representa de manera adecuada a las variables observadas.

Para explicar el uso que tienen estos métodos en psicometría y medición educacional primero vamos a llamar constructo a la propiedad o atributo teórico que se quiere medir. Por ejemplo en el caso de la prueba CAM el constructo que se pretende medir es el *conocimiento de alumnos y matemática*. Generalmente los constructos no pueden observarse directamente y tienen que ser medidos a través de otras variables, como por ejemplo los ítems de una prueba o preguntas de una encuesta.

En este contexto los análisis factoriales exploratorio y confirmatorio permiten [44]: determinar la validez de medición, esto se trata de determinar si los factores que explican los datos coinciden con los constructos teóricos que se pretenden medir; desarrollar teorías respecto a la naturaleza de los constructos y resumir un conjunto de variables en un número menor de factores que se puedan utilizar en otros análisis.

En este capítulo se describen los métodos de análisis factorial usual, análisis de componentes principales y análisis factorial confirmatorio y su aplicación sobre los datos de la prueba CAM. La aplicación se hizo con el objetivo de:

- Investigar la prueba CAM y analizar que dimensiones subyacentes está midiendo
- Seleccionar ítems que tengan buena calidad técnica y que permitan lograr una prueba con validez.
- Luego de alcanzar la validez de medición, relacionar el constructo *conocimiento de alumnos y matemática* con el *conocimiento matemático común y específico* medido por la prueba LMT. En particular se va a testear la hipótesis de que ambos constructos están correlacionados pero que se pueden distinguir.

2.1. Análisis Factorial

El modelo de análisis factorial busca representar un grupo de variables observadas x_i $i = 1, \dots, p$ a través de otro conjunto de variables llamadas factores por medio de un modelo lineal. Estos factores suelen interpretarse como variables latentes, es decir como variables subyacentes que no pueden medirse directamente, pero que se miden a través de las variables observadas.

Se supone que las variables observadas están estandarizadas, esto es que $\mathbb{E}(x_i) = 0$ y que $\text{var}(x_i) = \sigma_i^2 = 1$ (en el caso que no lo estén, se pueden estandarizar restándoles su media y dividiéndolas por su desviación estándar). Los factores que representen a las variables pueden ser de dos tipos:

- Factores comunes presentes en todas las variables, éstos se denotan como f_j con $j = 1, \dots, m$, donde $m \leq p$.
- Factores específicos que forman parte de una sola variable, los que se denotan como u_k con $k = 1, \dots, p$

La idea del modelo es explicar toda la varianza en común presente en las variables en término de los factores f y que los factores u sólo den cuenta de la varianza específica de cada variable. Para poder tener factores comunes es necesario que las variables tengan correlaciones no nulas entre sí, sino el único conjunto suficiente para explicar las variables serían ellas mismas [30].

El modelo matemático se define como:

$$x_i = \lambda_{i1}f_1 + \dots + \lambda_{im}f_m + L_{ii}u_i \quad (2.1)$$

De donde se puede distinguir la parte de la variable que es explicada por los factores comunes, que también explican al resto de las variables, y su parte específica:

$$x_i = \underbrace{\lambda_{i1}f_1 + \dots + \lambda_{im}f_m}_{\text{Factores comunes}} + \underbrace{L_{ii}u_i}_{\text{Factor específico de la variable } x_i} \quad (2.2)$$

Donde λ_{ij} son los coeficientes de la matriz Λ llamada matriz de cargas, ésta relaciona a los factores comunes con las variables observadas. La matriz L contiene los coeficientes L_{ii} que relacionan a los factores específicos con las variables observadas. Como cada factor específico forma parte de sólo una variable observada, la matriz L es diagonal. Por último denotando a $x = (x_1, \dots, x_p)$, $f = (f_1, \dots, f_m)$ y $u = (u_1, \dots, u_p)$ el modelo factorial resulta en:

$$x = \Lambda f + Lu \quad (2.3)$$

Para poder determinar los parámetros del modelo, se supone que todos los factores (comunes y específicos) tienen esperanza nula y varianza igual a 1, esto es:

$$\begin{aligned} \mathbb{E}(f_j) &= 0 & \forall j &= 1, \dots, m \\ \mathbb{E}(u_k) &= 0 & \forall k &= 1, \dots, p \\ \mathbb{E}(f_j^2) &= 1 & \forall j &= 1, \dots, m \\ \mathbb{E}(u_k^2) &= 1 & \forall k &= 1, \dots, p \end{aligned} \quad (2.4)$$

Como los factores comunes y específicos tienen que representar componentes independientes de las variables observadas, el modelo también supone que las correlaciones entre ellos son nulas.

$$\mathbb{E}(f_j u_k) = 0 \quad \forall j = 1 \dots m \text{ y } \forall k = 1, \dots, p \quad (2.5)$$

Además, como cada factor u_k sólo tiene que representar a la variable x_k , y se supone que toda la varianza común es descrita por los factores comunes, se supone que los factores específicos son incorrelados entre sí, esto es:

$$\mathbb{E}(u_k u_l) = 0 \quad \forall k = 1 \dots p \text{ y } \forall l = 1, \dots, p, \quad l \neq k \quad (2.6)$$

2.1.1. Descomposición de la varianza de las variables observadas

Considerando que las variables están estandarizadas y usando los supuestos (2.4),(2.5) y (2.6) la varianza de cada variable es:

$$1 = \sigma_i^2 = \sum_{k=1}^m \sum_{l=1}^m \lambda_{ik} \lambda_{il} \mathbb{E}(f_k f_l) + L_{ii}^2 \quad (2.7)$$

Por la indeterminación del modelo se puede decidir si los factores comunes son o no correlacionados. En el caso que se determine que tienen correlaciones nulas entre sí se tiene que

$$1 = \sigma_i^2 = \sum_{k=1}^m \lambda_{ik}^2 + L_{ii}^2. \quad (2.8)$$

A partir de la ecuación (2.7) se define la comunalidad h_i^2 de la variable x_i como el porcentaje de la varianza de x_i que es explicada por los factores comunes f , su expresión algebraica es:

$$h_i^2 = \sum_{k=1}^m \sum_{l=1}^m \lambda_{ik} \lambda_{il} \mathbb{E}(f_k f_l), \quad (2.9)$$

en el caso en que los factores comunes son incorrelados se tiene que:

$$h_i^2 = \sum_{k=1}^m \lambda_{ik}^2. \quad (2.10)$$

Al término L_{ii}^2 se le llama varianza específica y es la varianza de la variable x_i que no es explicada por los factores comunes. Algunos autores consideran que esa varianza está formada por dos componentes, una varianza verdadera s_i^2 dada por fluctuaciones propias de la variable x_i , y otra parte e_i^2 dada por las variaciones del error de medición de la variable. Se llama error de medición al error que se observa cuando una variable medida varias veces sobre el mismo objeto no entrega medidas idénticas, la varianza de este error es e_i^2 .

$$L_{ii}^2 = s_i^2 + e_i^2 \quad (2.11)$$

Se tiene entonces que:

$$h_i^2 = 1 - L_{ii}^2 \quad (2.12)$$

2.1.2. Relación entre las correlaciones de las variables y los parámetros del modelo

A partir de las ecuaciones (2.3),(2.4),(2.5),(2.6) y de que las variables observadas están estandarizadas se tiene que la correlación entre dos variables observadas se puede escribir como¹

$$r_{x_i x_k} = \mathbb{E}(x_i x_k) = \sum_{l=1}^m \sum_{q=1}^m \lambda_{il} \lambda_{kq} \mathbb{E}(f_l f_q) + L_{ii} L_{kk} \delta_{ik}$$

En el caso en que los factores comunes son incorrelados se tiene que:

$$r_{x_i x_k} = \sum_{l=1}^m \lambda_{il} \lambda_{kl} + L_{ii} L_{kk} \delta_{ik} \quad (2.13)$$

Si definimos $R_{xx} = \mathbb{E}(xx')$ a la matriz de correlaciones de las variables observadas y R_{ff} a la matriz de correlaciones de los factores comunes, tenemos que matricialmente las ecuaciones de correlaciones y varianza de las variables observadas (2.13) y (2.8) se escriben como

$$R_{xx} = \Lambda R_{ff} \Lambda' + L^2 \quad (2.14)$$

Esta ecuación se puede reescribir como:

$$R_{xx} - L^2 = \Lambda R_{ff} \Lambda' \quad (2.15)$$

En el caso de que los factores comunes sean incorrelados y de varianza igual a 1, (2.15) se escribe como:

$$R_{xx} - L^2 = \Lambda \Lambda' \quad (2.16)$$

Usando la ecuación (2.12), se tiene que la matriz $R_{xx} - L^2$ contiene en su diagonal las comunalidades h_i^2 de cada variable observada x_i . Es decir

$$R_{xx} - L^2 = \begin{pmatrix} h_1^2 & r_{x_1 x_2} & \cdots \\ r_{x_2 x_1} & h_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.17)$$

Para caracterizar el modelo es necesario estimar los siguientes parámetros: Λ , L y R_{ff} . La estimación de estos parámetros tendrá como objetivo reproducir las correlaciones y comunalidades de las variables. Es por esto que la estimación de la matriz Λ es la que más aproxime $\Lambda R_{ff} \Lambda'$ a la matriz $R_{xx} - L^2$. La matriz L^2 contiene a las varianzas específicas de cada variable, estimar esta matriz es equivalente a estimar las comunalidades gracias a la ecuación (2.12). Sin embargo el problema de estimar las comunalidades no se resuelve de manera única y dentro de los algoritmos de estimación se verá que existen diversas opciones para estimarlas, en particular se verá que el Análisis de Componentes Principales se puede entender como un análisis factorial donde las comunalidades se estiman como idénticas a 1.

Además se define la matriz de estructura como la matriz de correlaciones entre las variables observadas x y los factores comunes f . Se denota como R_{xf} y es de dimensión $p \times m$. Usando

¹ δ_{ik} se refiere al delta de Kronecker, que vale 0 si $k \neq i$, y 1 si $k = i$

la ecuación (2.3) se tiene que la correlación entre una variable observada x_i y un factor f_j es:

$$r_{x_i f_j} = \sum_{k=1}^m \lambda_{ik} \mathbb{E}(f_k f_j) = (\Lambda R_{ff})_{ij}$$

Luego, si se impone que los factores comunes sean incorrelados, la matriz de estructuras coincide con la matriz de cargas. Es decir los coeficientes de la matriz de cargas son correlaciones entre las variables observadas y los factores comunes. Esto es:

$$\Lambda = R_{xf} \tag{2.18}$$

2.1.3. Identificación del modelo

Para que el modelo esté identificado es necesario que el número de incógnitas sea menor o igual al número de ecuaciones. En el caso en que estas cantidades son iguales el modelo se podría calcular exactamente, en el caso en que el número de incógnitas es menor al de ecuaciones, el modelo se puede aproximar. Esto limita el número de factores que se pueden extraer en relación al total de variables observadas. La cota que representa la relación que tiene que cumplirse entre el número de variables observadas p y el número de factores comunes m es la siguiente [32]:

$$(p - m)^2 \geq p + m \tag{2.19}$$

Esta cota es una condición necesaria para que el modelo pueda estimarse.

Existe además otra indeterminación del modelo, respecto a la rotación de sus factores. Esto se observa considerando la ecuación principal del modelo:

$$x = \Lambda f + Lu \tag{2.20}$$

Si se define $\Lambda^* = \Lambda P$ y $f^* = P^{-1} f$ donde P es una matriz invertible de dimensión $m \times m$ entonces se cumple que:

$$x = \Lambda^* f^* + Lu \tag{2.21}$$

Esto implica que el modelo es invariante bajo rotaciones y que se pueden elegir factores comunes correlados o incorrelados que representen igual de bien a los datos. También esta indeterminación permite buscar factores que cumplan con alguna estructura específica.

2.1.4. Estimación de las comunalidades y la matriz de cargas

Para describir los métodos de estimación se supone que se sabe el número m de factores a extraer. Los métodos para escoger este número están descritos en la subsección 2.1.7. Vamos a partir describiendo cómo se estiman las comunalidades y luego como se aplican los tres métodos mas importantes para estimar la matriz de cargas, estos son Ejes Principales (EP), Máxima Verosimilitud (MV) y Residuos Mínimos (MinRes).

Para describir como se estima la comunalidad de una variable, se define primero al coeficiente de correlación múltiple:

Definición 2.1 (Coeficiente de correlación múltiple) *Sea Y la recta definida por la regresión lineal hecha entre la variable x_i y el resto de las variables $(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$. El coeficiente de correlación múltiple entre x_i y $(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ es:*

$$R_{i \bullet 1, 2, \dots, i-1, i+1, \dots, p}^2 = \text{cor}(x_i, Y) \quad (2.22)$$

La estimación de las communalidades se basa en una cota que relaciona la comunalidad de la variable observada i : h_i^2 , con el cuadrado del coeficiente de correlación múltiple entre esta variable y el resto.

$$R_{i \bullet 1, 2, \dots, i-1, i+1, \dots, p}^2 \leq h_i^2 \leq 1 \quad (2.23)$$

La cota superior se tiene por la ecuación (2.12). La cota inferior se tiene por el siguiente teorema y su corolario [30]:

Teorema 2.2 *El cuadrado del coeficiente de correlación múltiple para predecir la variable i a partir de las $p - 1$ variables restantes y los m factores comunes es igual a la comunalidad de la variable.*

DEMOSTRACIÓN. Véase [30] pp. 146-147 □

A partir del teorema anterior se puede inferir el corolario siguiente usando el hecho de que la correlación múltiple entre una variable y un conjunto de variables es siempre mayor a la correlación múltiple entre esa variable y un subconjunto del conjunto original.

Corolario 2.3 *El cuadrado del coeficiente de correlación múltiple para predecir la variable i a partir de las $p - 1$ variables restantes es una cota inferior para la comunalidad de la variable.*

Observación Debido al Corolario 2.3, muchos softwares utilizan como estimación inicial de la comunalidad de una variable observada al coeficiente de correlación múltiple entre esa variable y el resto de las variables.

Como se dijo anteriormente existe mucha indeterminación en los modelos y es por esto que la estimación calcula factores incorrelados. Esto implica que en todos los métodos que se describirán se tiene la siguiente ecuación para la matriz de correlaciones, la matriz de cargas y de varianzas específicas:

$$R_{xx} = \Lambda\Lambda + L^2 \quad (2.24)$$

Para obtener factores correlados hay que primero extraer los factores incorrelados y después aplicar una rotación.

Ejes Principales

El enfoque de ejes principales busca encontrar los factores que tengan máxima varianza y que sean incorrelados entre sí. Este enfoque es muy similar al de componentes principales, la mayor diferencia es que en ejes principales se buscan factores que expliquen sólo la varianza

común mientras que en componentes principales explican la varianza total. Este tema será discutido con mayor detalle en la sección de análisis de componentes principales.

Vamos a usar el estimador de la matriz R_{xx} dado por máxima verosimilitud S . Si llamamos $\hat{\Lambda}$ al estimador de Λ y \widehat{L}^2 al estimador de la matriz de varianzas específicas, es necesario que se cumpla la siguiente ecuación:

$$S - \widehat{L}^2 = \hat{\Lambda}\hat{\Lambda}' \quad (2.25)$$

Vamos a suponer que partimos con una estimación inicial de \widehat{L}^2 , la cual es generalmente (debido a la cota dada por (2.23)):

$$\widehat{L}^2 = \begin{pmatrix} 1 - \hat{h}_1^2 & 0 & \dots \\ 0 & 1 - \hat{h}_2^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 1 - R_{1\bullet 2, \dots, p}^2 & 0 & \dots \\ 0 & 1 - R_{2\bullet 1, 3, \dots, p}^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.26)$$

También puede utilizarse $\widehat{L}^2 = 0$. Como la matriz $S - \widehat{L}^2$ es simétrica, se puede diagonalizar, luego si llamamos P y D a las matrices que diagonalizan $S - \widehat{L}^2$ se tiene que:

$$S - \widehat{L}^2 = PDP' \quad (2.27)$$

Se supone que el m está elegido de forma que con m factores se debiera de poder representar completamente la varianza en común de las variables observadas que está reproducida en la matriz $S - \widehat{L}^2$. Luego se debiera de poder separar a la matriz P y D en las siguientes submatrices:

$$D = \begin{pmatrix} D_{m \times m}^1 & 0_{m \times (p-m)} \\ 0_{(p-m) \times m} & D_{(p-m) \times (p-m)}^2 \end{pmatrix} \quad (2.28)$$

$$P = \begin{pmatrix} P_{p \times m}^1 & P_{p \times (p-m)}^2 \end{pmatrix} \quad (2.29)$$

En donde $D^2 \approx 0$ (si no es así, se debiera aumentar el número de factores m). Con esto el método se resume en el siguiente algoritmo [32]:

1. Diagonalizar $S - \widehat{L}_i^2$ con la matriz de vectores propios P_i y con la matriz de valores propios D_i . Esto es:

$$S - \widehat{L}_i^2 = P_i D_i P_i' \quad (2.30)$$

Definir $\hat{\Lambda}_i = P_i^1 (D_i^1)^{\frac{1}{2}}$ donde D_i^1 contiene los m mayores valores propios, y P_i^1 a los vectores propios asociados. Si $\|\hat{\Lambda}_i^i - \hat{\Lambda}_i^{i-1}\| \leq \varepsilon$ parar el proceso, si no continuar.

2. Definir $\widehat{L}^{i+1} = \text{diag}(S - \hat{\Lambda}_i \hat{\Lambda}_i')$ y volver al punto 1

Máxima Verosimilitud

El enfoque de máxima verosimilitud supone que las variables observadas y los factores se distribuyen de forma normal multivariada. Si consideramos que x_1, \dots, x_N es una muestra

de N realizaciones independientes de la variable estandarizada x , donde cada realización x_i es de dimensión $p \times 1$ la función de verosimilitud para datos normales multivariados es [32]:

$$f(x_1, \dots, x_N | \mu, R_{xx}) = \prod_{i=1}^N |R_{xx}|^{-\frac{1}{2}} (2\pi)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2} (x_i - \mu)' R_{xx}^{-1} (x_i - \mu) \right\} \quad (2.31)$$

Tomando el logaritmo de esta función y utilizando que $\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})' = S$ más algunos manejos algebraicos, la verosimilitud resulta en (sin considerar las constantes que dependen de p o de N) [32]:

$$L(\mu, R_{xx} | x_1, \dots, x_N) = -\frac{N}{2} \log(|R_{xx}|) - \frac{N}{2} \text{tr}(R_{xx}^{-1} S) - \frac{N}{2} (\bar{x} - \mu)' R_{xx}^{-1} (\bar{x} - \mu) \quad (2.32)$$

Remplazando μ por el estimador promedio de la muestra \bar{x} resulta:

$$L(R_{xx} | x_1, \dots, x_N) = -\frac{N}{2} \log(|R_{xx}|) - \frac{N}{2} \text{tr}(R_{xx}^{-1} S) \quad (2.33)$$

Luego, usando que $R_{xx} = \Lambda\Lambda + L^2$, la función a maximizar resulta en:

$$L(\Lambda, L^2) = -\frac{N}{2} \log(|\Lambda\Lambda + L^2|) - \frac{N}{2} \text{tr}((\Lambda\Lambda + L^2)^{-1} S) \quad (2.34)$$

El proceso de maximización de esta función parte con una estimación inicial de L^2 (que puede ser dada por ejes principales o los coeficientes de correlación múltiple), luego se estima Λ , y se vuelve a estimar L^2 . Este proceso continúa iterativamente hasta que la diferencia entre estimaciones sea menor que una tolerancia dada (es decir $\|\hat{\Lambda}_{i+1} - \hat{\Lambda}_i\| < \varepsilon$ y $\|\hat{L}_{i+1}^2 - \hat{L}_i^2\| < \varepsilon$). Para determinar la solución se impone que $\hat{\Lambda}'(\hat{L}^2)^{-1}\hat{\Lambda} = D$ con D diagonal.

Residuos Mínimos (MinRes)

En este enfoque lo que se busca es ajustar sólo los elementos de fuera de la diagonal de la matriz de correlaciones. Esto se realiza considerando que sólo esos representan estrictamente relaciones de las variables. Es por esto que la función que minimizan es [25]:

$$f(R_{xx}, \Lambda) = \sum_{i=2}^p \sum_{j=1}^{i-1} (R_{xx} - \Lambda\Lambda')_{ij}^2 \quad (2.35)$$

Lo interesante de este método es que no necesita estimar las comunalidades y que sin embargo al final entrega una estimación de ellas. Es más, se considera que está relacionado con el método de ejes principales porque si se utiliza el método de ejes principales con las comunalidades estimadas por MinRes el resultado es el mismo de MinRes [13].

Observación En los tres métodos descritos durante el proceso de estimación, se produce una descomposición en valores y vectores propios en donde los valores propios se interpretan como la varianza explicada por cada factor (ver [32] y [13]). Siempre se considera como el primer factor al relacionado con el mayor valor propio.

2.1.5. Elección del método

Primero hay que considerar que hay dos casos en los cuales la elección del método no es muy relevante ya que se obtienen resultados similares con todos ellos [13], estos son :

- Cuando las comunalidades son cercanas a 1, el uso de ejes principales, máxima verosimilitud, MinRes o componentes principales entregan aproximadamente la misma solución.
- Cuando hay muchas variables, los elementos de la diagonal son menos significativos, ya que en la diagonal hay p elementos y sobre la diagonal hay $(p^2 - p)/2$ elementos (que son los mismos que están bajo la diagonal, ya que la matriz es simétrica). Luego la estimación de las comunalidades tiene menos importancia y los métodos dan resultados similares entre sí.

Hay que considerar especialmente al método de estimación con máxima verosimilitud, ya que sus estimadores son eficientes, consistentes y permiten testear hipótesis [4]. Por otra parte la estimación con máxima verosimilitud es mas pesada computacionalmente que los otros métodos y necesita que la matriz estimada de correlaciones de las variables observadas sea no singular. Luego en el caso en que la matriz de correlaciones estimada sea cercana a una matriz singular o que el método falle en converger es preferible utilizar los otros métodos [33] (esto puede ocurrir por ejemplo cuando se usa la matriz de correlaciones tetracóricas).

También es importante considerar que los métodos de ejes principales y de residuos mínimos no realizan supuestos sobre la distribución de las variables, luego en casos de falla de la hipótesis de normalidad puede resultar mejor utilizar estos métodos.

2.1.6. Rotaciones

Como el modelo de análisis factorial es indeterminado y los métodos consideran ciertos supuestos arbitrarios para identificarlo y poder llegar a una solución, se puede cambiar la forma de los factores para obtener una solución más fácil de interpretar. Esto se hace mediante una rotación la cual no afecta la calidad del ajuste del modelo (ver las ecuaciones (2.20) y (2.21)).

El autor que propuso inicialmente este objetivo fue Thurstone, el cual planteó el concepto de estructura simple que es que cada variable sea representada por un número menor de factores que el número de factores comunes total. Este concepto se materializa imponiendo que la matriz de estructura (que coincide con la matriz de cargas cuando los factores comunes son incorrelados) tenga al menos un 0 en cada fila, es decir que cada variable sea incorrelada o no esté influida por al menos un factor [30].

Hay diversas formas de rotar los factores para lograr una estructura simple, algunas entregan factores incorrelados y otras correlados. En particular en psicometría y medición educacional los constructos que se investigan suelen ser correlados por lo tanto generalmente se proponen rotaciones que generan factores correlados para facilitar la interpretación.

Las rotaciones que se presentan son Varimax, que genera factores con estructura simple e incorrelados y Promax que genera factores con estructura simple y permite que estén correlados entre sí. Estas rotaciones son comunmente usadas, pero existen varias más, las que

no se describen porque no fueron utilizadas.

Varimax

Supongamos que se tiene la matriz Λ , la cual se rota con una matriz T ortogonal para obtener a la matriz de cargas $\Lambda^* = \Lambda T$ con el objetivo de que alcance una estructura simple. El tipo de estructura simple que se busca es que los factores influyan en grupos separados de variables, que permitan darle una interpretación significativa a cada factor. Para esto se define a la varianza de los cuadrados de los coeficientes de la columna j de la matriz Λ^* :

$$v_j = \frac{1}{p} \sum_{i=1}^p (\lambda_{ij}^{*2}) - \frac{1}{p} \sum_{k=1}^p \lambda_{kj}^{*2} \quad (2.36)$$

Para tomar en cuenta a todos los factores estas varianzas se suman en el siguiente término:

$$v = \sum_{j=1}^m \frac{1}{p} \left(\sum_{i=1}^p (\lambda_{ij}^{*2})^2 - \frac{1}{p} (\sum_{k=1}^p \lambda_{kj}^{*2})^2 \right) \quad (2.37)$$

Para que el término v se maximice, cada factor tiene que tener correlaciones altas en valor absoluto con las variables y correlaciones cercanas a cero. Esto proporciona una matriz de estructura o de cargas con estructura simple. Por lo tanto para encontrar la matriz T se maximiza:

$$\max_{T \text{ ortogonal}} v(T) \quad (2.38)$$

Existe una variación del método Varimax, llamado Varimax Normalizado donde se divide a los coeficientes de la matriz por la comunalidad de cada variable. En este enfoque la función a maximizar es:

$$\tilde{v}(T) = \sum_{j=1}^m \left(\frac{1}{p} \sum_{i=1}^p (\lambda_{ij}^{*2}/h_i^2)^2 - \frac{1}{p^2} \left(\sum_{k=1}^p \lambda_{kj}^{*2} h_p^2 \right)^2 \right) \quad (2.39)$$

El método Varimax Normalizado funciona mejor que el Varimax simple y entrega soluciones que se consideran muy buenas [30].

Promax

El método de rotación Promax se basa en un método más general, que busca una matriz de rotación T invertible, pero no necesariamente ortogonal, que logre que la matriz $\Lambda^* = \Lambda T$ sea lo más similar posible a una matriz objetivo B . Es decir se quiere que:

$$\Lambda T \approx B \quad (2.40)$$

Consideremos que se realiza un análisis factorial y luego la matriz de cargas se transforma por medio de una rotación Varimax en la matriz Λ^+ . Lo que hace el método Promax es buscar una matriz T invertible que acerque a la matriz de cargas original Λ a una transformación de Λ^+ . Es decir $B = f(\Lambda^+)$ donde la transformación f lo que hace es [30]:

Primero normaliza las filas de la matriz Λ^+ con la norma euclidiana:

$$B_{ij}^0 = \frac{\lambda_{ij}^+}{\sum_{k=1}^m \lambda_{ik}^+}$$

Luego normaliza las columnas con la norma 1:

$$B_{ij}^1 = \frac{B_{ij}^0}{\|B_{\bullet j}^0\|_1}$$

Luego eleva los coeficientes de B^1 por un número κ , para que se acentúe la cercanía a cero de los que son casi nulos, y para que se acentúe la lejanía a cero de los que están lejos. Se recomienda usar $\kappa=4$ [30]. Se utiliza el módulo para mantener el signo original del coeficiente:

$$B_{ij} = \frac{|B_{ij}^1|^{\kappa+1}}{B_{ij}^1}$$

El método Promax logra soluciones similares a Varimax, pero con factores correlacionados, que suelen tener una estructura simple.

2.1.7. Elección del número de factores o componentes

Determinar el número de factores en el análisis factorial es una decisión crítica, ya que la estimación de los parámetros depende del número de factores escogidos (en el caso de componentes principales no existe esa dependencia). Hay muchas estrategias posibles, alguna comunes porque suelen ser la opción por defecto que utilizan los softwares. Lo ideal es usar varios criterios que ojalá coincidan, aunque al final la decisión depende del juicio de la persona [44]. Estas estrategias también sirven para determinar el número de factores a extraer cuando se usa análisis de componentes principales.

Valores propios mayores a 1

En todos los métodos de estimación se calculan valores propios que representan la varianza explicada por cada factor común. Como se trabaja bajo el supuesto de que las variables están estandarizadas, es decir que sus varianzas son 1. Un factor que está asociado a un valor propio menor que 1 estaría explicando menos varianza que una variable observada aislada. Bajo esta lógica, que fue propuesta por Guttman y Kayser, el total de factores a extraer serían los que explican más varianza que una variable aislada, es decir el total de valores propios mayor a 1. El problema de este método es que es muy sensible a errores de cálculo o de muestreo.

Scree Test

Este procedimiento considera el gráfico de sedimentación, el cual se construye graficando la varianza explicada por cada factor, es decir graficando los valores propios que entrega el análisis. Este procedimiento fue propuesto por Raymond Cattell [8], que buscó un criterio para determinar el número de factores de manera simple y significativa. Lo que propone el test es identificar en el gráfico de sedimentación donde se producen aumentos sustantivos de la varianza, estos formarían un “codo”. Luego el criterio es elegir tantos factores como valores propios haya antes del codo. En el caso en que se formen varios codos, se usa como referencia

el último de ellos.

Como este criterio se basa en la observación de un gráfico es un poco subjetivo y es recomendable apoyarlo con otras técnicas.

Análisis paralelo

Lo que propone este análisis es simular variables con la misma distribución que las originales, pero incorreladas entre sí. Esta simulación genera una muestra del mismo tamaño que la original, a la cual se le aplica un análisis factorial o análisis de componentes principales y se analizan sus valores propios, que representan sólo varianza debida a error aleatorio. Luego el comparar estos valores propios con los del análisis sobre las variables originales permite tener una noción respecto de si los valores propios toman en cuenta más que varianza dada por error de muestreo.

Hay muchas formas de concretizar esta idea, en particular en esta memoria se tomó la muestra de cada variable y se reordenó de manera aleatoria. Con estas variables ordenadas de manera aleatoria se realizaron análisis factoriales, o de componentes principales, según corresponda.

2.2. Análisis de Componentes Principales

El objetivo principal del análisis de componentes principales ACP es representar p variables observadas en un espacio de dimensión m con $m \leq p$ conservando la máxima variabilidad posible de los datos. Para lograr esto se transforman las p variables observadas en m variables, a través de una transformación lineal, que logre que las nuevas variables tengan una varianza máxima y que sean ortogonales entre sí [23]. A las m nuevas variables se les llama componentes principales de las variables observadas.

Las componentes cumplen con que la primera componente principal es la combinación lineal de variables observadas con máxima varianza, la segunda componente principal es la combinación lineal de las variables observadas con máxima varianza e incorrelada con la primera componente principal. En general la k -ésima componente principal va a ser la combinación lineal de las variables observadas con máxima varianza y que cumpla con ser incorrelada con la primera hasta la $(k - 1)$ -ésima componente. Por esto la varianza explicada por cada componente es decreciente, y la que explica la mayor cantidad posible de varianza es la primera. Esto permite usar el análisis de componentes principales cuando se necesite crear un puntaje que resuma un grupo de variables. En el caso de una prueba donde los ítems son las variables observadas esta propiedad es muy importante ya que permite asignarle un puntaje a cada persona con una mayor riqueza interpretativa que la suma o promedio.

Este método es también un primer paso para identificar posibles variables latentes o no observadas que están generando variabilidad en los datos [32] y al generar variables incorreladas, permite distinguir e identificar factores independientes que explican los datos.

Notación:

- Llamaremos X a la matriz con los datos de las p variables observadas sobre los N individuos. La dimensión de la matriz es $N \times p$ donde

$$X_{ij} = \text{Valor de la variable } j \text{ en el individuo } i$$

Vamos a suponer que a cada variable se le restó su media. Es decir si la variable original era \tilde{x}_j , donde $\tilde{x}_j(i)$ es el valor de la variable en el individuo i , la variable que se utilizará es

$$x_j(i) = \tilde{x}_j(i) - \frac{1}{N} \sum_{k=1}^N \tilde{x}_j(k)$$

- Llamaremos S a la matriz de covarianzas estimadas, como a las variables se les restó su media la matriz S se calcula como $S = \frac{1}{N} X'X$.

2.2.1. Componentes principales como combinaciones lineales de los datos que tienen máxima varianza

Vamos a definir a las componentes principales como las combinaciones lineales de las variables originales con máxima varianza que además son incorreladas entre sí. Consideremos

a los vectores Λ^1, Λ^2 hasta Λ^p , cada uno de dimensión $p \times 1$. Estos vectores van a definir las componentes principales de la siguiente forma:

$$F^i = X\Lambda^i \quad i = 1, \dots, p \quad (2.41)$$

La solución a este problema es que los vectores Λ^i sean los vectores propios de la matriz $X'X$ asociados a los valores propios ordenados de forma decreciente.

Teorema 2.4 *Si la matriz $X'X$ es no singular, se pueden encontrar las p componentes principales de la matriz de variables observadas X . Las cuales se definen como*

$$F^k = X\Lambda^k \quad \text{con } k = 1, \dots, p \quad (2.42)$$

Donde Λ^k es el vector propio asociado al k -ésimo mayor valor propio δ_k de la matriz $X'X$. Además se tiene que $X = F\Lambda'$ donde $\Lambda = [\Lambda^1 \dots \Lambda^p]$ es la matriz de dimensión $p \times p$ formada por los vectores propios de $X'X$ y $F = [F^1 F^2 \dots F^p] = X\Lambda$ es una matriz de dimensión $N \times p$ formada por las componentes principales de X .

DEMOSTRACIÓN. **Primer Caso:** $F^1 = X\Lambda^1$

Tenemos que maximizar $\text{var}(F^1) = \frac{1}{N}F^{1'}F^1$, esto es lo mismo que maximizar $F^{1'}F^1 = \Lambda^{1'}X'X\Lambda^1$. A esta optimización hay que agregarle la restricción: $\Lambda^{1'}\Lambda^1 = 1$, ya que sin esta condición sería imposible llegar a un máximo porque la varianza podría crecer indefinidamente. El lagrangiano de este problema es:

$$L = \Lambda^{1'}X'X\Lambda^1 - \delta_1(\Lambda^{1'}\Lambda^1 - 1) \quad (2.43)$$

Derivando e igualando a 0 se tiene que:

$$\begin{aligned} \frac{dL}{d\Lambda^1} &= 2X'X\Lambda^1 - 2\delta_1\Lambda^1 \\ &= 0 \end{aligned}$$

De esto se obtiene $X'X\Lambda^1 = \delta_1\Lambda^1$. Luego Λ^1 es un vector propio de la matriz $X'X$ asociado al valor propio δ_1 , esto implica que $\text{var}(F^1) = \frac{1}{N}\Lambda^{1'}X'X\Lambda^1 = \frac{1}{N}\Lambda^{1'}\delta_1\Lambda^1 = \frac{1}{N}\delta_1$. Luego para que F^1 tenga varianza máxima Λ^1 tiene que estar asociado al mayor valor propio.

k-1 \Rightarrow k: En este caso, al igual que en el primero $\text{var}(F^k) = \frac{1}{N}F^{k'}F^k = \Lambda^{k'}X'X\Lambda^k$, sin embargo aparte de la restricción $\Lambda^{k'}\Lambda^k = 1$ es necesario agregar que la componente F^k debe ser incorrelada con las componentes anteriores, es por esto que el lagrangiano resulta en:

$$L = \Lambda^{k'}X'X\Lambda^k - \delta_k(\Lambda^{k'}\Lambda^k - 1) + \sum_{i=1}^{k-1} \phi_i \Lambda^{k'}\Lambda^i \quad (2.44)$$

derivando e igualando a 0 se tiene que:

$$\frac{dL}{d\Lambda^k} = 2X'X\Lambda^k - 2\delta_k\Lambda^k + \sum_{i=1}^{k-1} \phi_i \Lambda^i = 0 \quad (2.45)$$

Si premultiplicamos a $\frac{dL}{d\Lambda^k}$ por $\Lambda^{j'}$ con $j \leq k - 1$ y notamos que por hipótesis inductiva $\Lambda^{j'}$ es el vector propio asociado al j -ésimo valor propio tenemos que:

$$2\Lambda^{j'}X'X\Lambda^k - 2\delta_k\Lambda^{j'}\Lambda^k + \sum_{i=1}^{k-1} \phi_i\Lambda^{j'}\Lambda^i = 0 \quad (2.46)$$

$$\Rightarrow \phi_j\Lambda^{j'}\Lambda^j = 0 \quad (2.47)$$

$$\Rightarrow \phi_j = 0 \quad (2.48)$$

Esto porque por hipótesis inductiva, $\forall j \leq k - 1$ Λ^j es incorrelado con el primer hasta el k -ésimo vector propio y esto implica que:

$$\frac{dL}{d\Lambda^k} = 0 \Leftrightarrow X'X\Lambda^k = \delta_k\Lambda^k \quad (2.49)$$

Por lo tanto Λ^k es un vector propio. Como se quiere que f_k tenga máxima varianza Λ^k tiene que ser asociado al k -ésimo mayor valor propio, ya que $\text{var}(f_k) = \frac{1}{N}\delta_k$. Cómo Λ es la matriz de vectores propios normalizados, es una matriz ortonormal, es decir $\Lambda\Lambda' = I$. Luego como la matriz de factores se expresa como: $F = X\Lambda$, se tiene que $\Lambda F' = \Lambda\Lambda'X' = X'$. \square

Corolario 2.5 *La varianza de la componente k es el k -ésimo valor propio de la matriz S .*

DEMOSTRACIÓN. Como $S = \frac{1}{N}X'X$ su k -ésimo valor propio α_k es igual al k -ésimo valor propio de $X'X$ dividido por N . Luego $\text{var}(F^k) = \frac{1}{N}\delta_k = \alpha_k$. \square

Teorema 2.6 *Supongamos que se quieren calcular m componentes incorrelados con $1 \leq m \leq p$ cuya suma de varianzas sea máxima. Entonces una solución para este problema es tomar:*

$$[F_1 \dots F_m] = X[\Lambda^1 \dots \Lambda^m]$$

Donde $[\Lambda^1 \dots \Lambda^m]$ son los primeros m vectores propios de S . Es decir, si se extraen k componentes, estas no sólo maximizan su varianza individual con la restricción de ser incorreladas, si no que también maximizan la varianza total que representan.

DEMOSTRACIÓN. Véase [23] pp. 11-13. \square

Observación Como $S = \frac{1}{N}X'X$ se tiene que S es simétrica y semidefinida positiva (ver teorema 2.3 [4]). Por lo tanto, todos sus valores propios son no negativos. Esto es una propiedad teórica muy importante, ya que los valores propios de esta son la varianza de los respectivos componentes o factores. Por esto no tiene sentido que tomen valores negativos. Esta propiedad además es una ventaja que tiene el ACP sobre el análisis factorial usual, ya que en este último pueden aparecer valores propios que representan varianzas, negativos. A estos casos se les llama casos de Heywood y los softwares tratan de evitarlos o repararlos de diversas formas.

2.2.2. Relación entre el análisis factorial usual y el análisis de componentes principales

El modelo de análisis de componentes principales, lleva a la siguiente igualdad:

$$X = F\Lambda' \quad (2.50)$$

Con esta ecuación se hace evidente que se representan los datos observados (X) con las componentes (F). Esto es similar a lo que hace el análisis factorial, sólo que aquí se representan exactamente a los datos y se usan p componentes.

Sin embargo se puede entender al análisis de componentes principales, como un tipo de análisis factorial extrayendo en vez de las p componentes, sólo las primeras m . Para elegir este m también aplican los procedimientos de selección descritos en la subsección 2.1.7.

Para precisar como se puede ver el ACP como un análisis factorial se va a considerar la siguiente notación:

- A las componentes se les va a llamar factores.
- Se van a estandarizar los factores, es decir se van a dividir por su desviación estándar para que tengan varianza igual a 1. Llamando D a la matriz que tiene en su diagonal a los valores propios de S , se definen los factores estandarizados como:

$$\bar{F} = F \cdot D^{-\frac{1}{2}} \quad (2.51)$$

- Se define la matriz de cargas estandarizada $\bar{\Lambda}$ de dimensión $p \times p$ como:

$$\bar{\Lambda} = D^{\frac{1}{2}} \cdot \Lambda \quad (2.52)$$

- Se definen también a las siguientes matrices truncadas:

$\bar{F}_m = [\bar{F}^1 \ \bar{F}^2 \ \dots \ \bar{F}^m]$ es la matriz formada por las m primeras columnas de \bar{F}

$F_m = [F^1 \ F^2 \ \dots \ F^m]$ es la matriz formada por las m primeras columnas de F

$\bar{\Lambda}_m = [\bar{\Lambda}^1 \ \bar{\Lambda}^2 \ \dots \ \bar{\Lambda}^m]$ es la matriz formada por las primeras columnas de $\bar{\Lambda}$

$\Lambda_m = [\Lambda^1 \ \Lambda^2 \ \dots \ \Lambda^m]$ es la matriz formada por las primeras columnas de Λ

$D_m = D_{m \times m}$ es una matriz diagonal, que contiene los primeros m valores propios.

Con estas definiciones se tiene la siguiente igualdad:

$$X = \bar{F}\bar{\Lambda}' \quad (2.53)$$

Si se supone que la varianza de los datos puede explicarse bien con un número m de factores, con $m < p$, se puede generar un modelo similar al de análisis factorial:

$$X \approx \bar{F}_m \bar{\Lambda}'_m \quad (2.54)$$

Donde la matriz de covarianzas de \bar{F}_m es:

$$\text{cov}(\bar{F}_m) = \frac{1}{N} \bar{F}'_m \bar{F}_m = \frac{1}{N} D_m^{-\frac{1}{2}} \Lambda_m X' X \Lambda_m D_m^{-\frac{1}{2}} \quad (2.55)$$

$$= \begin{pmatrix} 1 & 0 & \dots \\ 0 & 1 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.56)$$

y por la ecuación (2.54)

$$S \approx \frac{1}{N} \bar{\Lambda}_m \bar{F}'_m \bar{F}_m \bar{\Lambda}'_m = \bar{\Lambda}_m \bar{\Lambda}'_m \quad (2.57)$$

La ecuación (2.57) permite explicitar la diferencia principal entre el modelo de análisis factorial usual y el análisis de componentes principales. En el análisis de componentes principales se busca una matriz $\bar{\Lambda}_m$ que se aproxime lo más posible a la matriz estimada de covarianzas S :

$$S \approx \bar{\Lambda}_m \bar{\Lambda}'_m$$

En el modelo de análisis factorial, si se imponen factores incorrelados, se busca una matriz $\bar{\Lambda}_m$ y una matriz L^2 tal que

$$S \approx \bar{\Lambda}_m \bar{\Lambda}'_m + L^2$$

En resumen, en el análisis factorial usual el objetivo es representar con los factores comunes f a la *varianza en común* que tienen las variables observadas. Es por esto que es necesario extraer de la varianza total la varianza específica, y dejar sólo la parte común, es decir las comunales de las variables y las correlaciones. En cambio en el análisis de componentes principales se busca explicar la *varianza total* de las variables observadas, es decir la común y la específica en conjunto [11].

Formalmente, si se mira el análisis de componentes principales a partir del enfoque del análisis factorial se tiene que:

- Las comunales se estiman como igual a 1 (es decir las varianzas únicas se estiman como 0).
- Se impone que los factores comunes sean incorrelados (i.e. $R_{ff} = I$).
- Para estimar la matriz Λ se usa el método de ejes principales sólo hasta la primera iteración.

Existe una discusión respecto a si es adecuado estimar el modelo de análisis factorial utilizando componentes principales, porque se estarían sobrestimando las comunales, sin embargo el análisis de componentes principales parte de otros objetivos a los del análisis factorial y es un modelo completo en sí mismo [10]. Sin embargo a pesar de que los modelos de análisis factorial y análisis de componentes principales tienen distintos objetivos los resultados que entregan sobre los datos son generalmente muy similares.

Cómo no importa cuantas componentes se consideren, ya que su cálculo sólo depende de la descomposición en valores y vectores propios de la matriz estimada de correlaciones S , el análisis de componentes principales no necesita el número m de factores comunes para determinar sus coeficientes. Esto no es así en el análisis factorial usual en donde los resultados dependen de ese número.

Otro punto importante, es que en el caso en que las variables no estén estandarizadas el análisis de componentes principales puede diferir mucho del análisis factorial usual, ya que da más peso a las variables con más varianzas. Por esto es recomendable usar el ACP con la matriz de correlaciones estimadas, más que con las de covarianzas (a menos que las varianzas tengan un significado relevante).

2.3. Análisis Factorial Confirmatorio

En el análisis factorial confirmatorio se propone una estructura para los factores y se prueba si esta estructura describe de manera adecuada a los datos. Esta técnica difiere del análisis factorial usual, donde la estructura de los factores es entregada por la estimación de los parámetros.

Se va a utilizar la siguiente notación:

- x es el vector de variables observadas, es de dimensión $p \times 1$ y se distribuye como $N_p(\mu, V)$
- f es un vector de factores o variables latentes de dimensión $m \times 1$ que se distribuye de manera normal, $f \sim N_m(0, I_{m \times m})$
- A es la matriz que define la relación entre x y f , es decir, es la matriz que entrega la estructura factorial. Puede tener coeficientes definidos como 0 o alguna otra constante. También puede tener coeficientes libres los que se estiman para que el modelo se ajuste a los datos. Al vector de coeficientes libres de esta matriz lo llamaremos θ , con lo cual $A = A(\theta)$
- ξ es el vector con perturbaciones no observadas, (similar al u del análisis factorial).
- P es la matriz de covarianzas del vector ξ , al igual que la matriz A puede tener coeficientes fijos y coeficientes variables que definen también la estructura del modelo propuesto. Los coeficientes variables son estimados de forma que maximicen el ajuste del modelo a los datos. Si llamamos β al vector de coeficientes variables de P , resulta que $P = P(\beta)$

Observación En el análisis factorial usual no se impone normalidad para las variables observadas (a menos que se utilice máxima verosimilitud para estimar). En el análisis factorial confirmatorio también se podría eludir este supuesto, sin embargo se mantiene porque la interpretación de muchos índices de ajuste dependen de él.

En el análisis factorial confirmatorio, se plantean las siguientes ecuaciones matriciales, la idea es que a través de ellas se defina la estructura factorial, dejando libres ciertos coeficientes y fijando otros [12]:

$$\begin{pmatrix} x \\ f \end{pmatrix} = A \begin{pmatrix} x \\ f \end{pmatrix} + \xi \quad (2.58)$$

$$\text{cov}(\xi) = P \quad (2.59)$$

Las ecuaciones (2.58) y (2.59) son el corazón de la estructura factorial que propone el investigador. En la matriz A es donde los coeficientes determinan que variables cargan en que factores y es en la matriz P donde se determina que factores están correlacionados entre sí.

Para entender como se ajusta el modelo, hay que escribir la matriz de covarianzas de x en función de los parámetros a estimar, para esto se define:

$$J = \begin{pmatrix} I_{p \times p} & 0 \\ 0 & 0 \end{pmatrix} \quad (2.60)$$

y utilizando la ecuación (2.58)

$$\begin{pmatrix} x \\ f \end{pmatrix} = (I - A)^{-1} \xi \quad (2.61)$$

Esto implica que teóricamente, el valor de la matriz de varianzas y covarianzas de x es:

$$\begin{aligned} V_{xx} = \Sigma &= \text{cov} \left(J \begin{pmatrix} x \\ f \end{pmatrix} \begin{pmatrix} x \\ f \end{pmatrix}' J' \right) \\ \Sigma &= J(I - A)^{-1} \text{cov}(\xi) (I - A)'^{-1} J' \\ \Sigma &= J(I - A)^{-1} P (I - A)' J' \end{aligned} \quad (2.62)$$

Definiendo al estimador de V_{xx} como:

$$\widehat{V}_{xx} = S \quad (2.63)$$

Luego, los parámetros de A y P se estiman de manera que se minimice algún tipo de distancia entre la matriz S de covarianzas empíricas y $\Sigma = J(I - A)^{-1} P (I - A)'^{-1} J'$. La distancia utilizada depende del método usado, en el caso del método de máxima verosimilitud la función a minimizar es la siguiente [31]:

$$L(\theta, \beta) = \text{tr}(S\Sigma^{-1}) - p + \ln(\det(\Sigma)) - \ln(\det(S)) \quad (2.64)$$

Con lo cual, para encontrar los estimadores de los parámetros θ y β , hay que resolver el problema no lineal

$$\min_{\theta, \beta} L(\theta, \beta) \quad (2.65)$$

Si $(\hat{\theta}, \hat{\beta})$ son la solución estimada de (2.65), entonces la matriz de covarianzas reproducida por el modelo es:

$$\hat{\Sigma}_0 = \Sigma(\hat{\theta}, \hat{\beta}) = J(I - A(\hat{\theta}, \hat{\beta}))^{-1} P(\hat{\theta}, \hat{\beta}) (I - A(\hat{\theta}, \hat{\beta}))'^{-1} J'$$

2.3.1. Notación utilizando grafos

Existe una notación muy práctica, que permite representar la estructura de la matriz A y P con grafos.

La notación con grafos consiste en que cada variable se representa con un nodo, y la relación entre las variables se representa por medio de flechas.

- Las variables observadas se representan con nodos de forma cuadrada
- Las variables latentes o factores se representan con nodos de forma circular u ovalada.
- Una flecha unidireccional \rightarrow representa que la variable en la cual incide la flecha, es influida por la variable desde la cual parte la flecha. Esta influencia significa que si llamamos $v_{inicial}$ a la variable desde la cual parte la flecha, y v_{final} a la variable que llega se tiene que

$$v_{final} = a_{j,k} \cdot v_{inicial} + \text{la influencia de otras variables}$$

donde $a_{j,k}$ es el coeficiente de la matriz A que relaciona a estas variables

Es importante notar que las flechas que no están presentes en el grafo, implican que los coeficientes asociados a ellas son 0, en particular si dos variables latentes no tienen una flecha bidireccional entre ellas, se está imponiendo que su correlación es 0.

2.3.2. Índices de ajuste del modelo

Existen en la literatura diversos índices que buscan dar cuenta de la calidad de un modelo de análisis factorial confirmatorio. Estos toman en cuenta la distancia entre la matriz de covarianzas estimadas S y la matriz de covarianzas reproducida por el modelo $\hat{\Sigma}_0$ y hay de tres tipos [35]:

- Índices que miden algún tipo de distancia entre S y $\hat{\Sigma}_0$. Ej: chi-cuadrado, RMSEA y SRMR.
- Índices que comparan el ajuste del modelo propuesto con el ajuste de un modelo base (que en general es el modelo nulo) Ej: GFI, AGFI.
- Índices que miden el ajuste entre S y $\hat{\Sigma}_0$ y también toman en cuenta la complejidad del modelo. Ej: Criterio de Akaike.

Coefficiente Chi-cuadrado χ^2

El coeficiente χ^2 mide la discrepancia entre la matriz de covarianzas muestral S y la matriz de covarianzas reproducida por el modelo $\hat{\Sigma}_0$ y su formulación depende del método de estimación usado.

En el caso de la estimación por máxima verosimilitud el estadístico de chi-cuadrado se calcula como:

$$\chi^2 = (N - 1) \cdot (tr(S\Sigma^{-1}) - p + \ln(\det(\Sigma)) - \ln(\det(S)))$$

donde N es el número de individuos en la muestra. Los grados de libertad de este estadístico son:

$$g.l. = \frac{p(p-1)}{2} - t$$

Donde $\frac{p(p-1)}{2}$ es el número de elementos no redundantes de S (ya que es una matriz simétrica) y t es el número de parámetros a estimar.

Como el objetivo del modelo es que la matriz de covarianzas muestral sea lo mas cercana a la matriz de covarianzas reproducida por el modelo, un buen ajuste se tendrá cuando el estadístico de chi-cuadrado sea pequeño, es decir cuando el p -valor sea grande y no se rechace que la matriz S es diferente a $\hat{\Sigma}_0$.

Sin embargo el estadístico de chi-cuadrado es muy dependiente del tamaño de la muestra, y para muestras grandes siempre va a rechazar la hipótesis $H_0 : \chi^2 = 0$, por lo tanto se recomienda también mirar la razón entre el estadístico χ^2 y sus grados de libertad $\frac{\chi^2}{g.l.}$, si esta razón es pequeña, se considera que hay un buen ajuste, una regla de la literatura es que esa razón debe ser menor que 2 para modelos buenos y entre 2 y 3 para modelos aceptables [35].

Error Cuadrático de Aproximación RMSEA

El error cuadrático medio de aproximación (en inglés Root Mean Square Error of Approximation RMSEA) permite tener un conocimiento del ajuste sobre los datos menos dependiente

del tamaño muestral que el chi-cuadrado. Es una medida global del ajuste del modelo. El estimador de este error se calcula cómo

$$RMSEA = \sqrt{\frac{\chi^2 - g.l.}{(N - 1) \cdot g.l.}}$$

En la literatura se considera que valores de este índice menores a 0,05 muestran un buen ajuste y entre 0,05 y 0,08 un ajuste regular [35].

Raíz de los cuadrados de los residuos estandarizados SRMR

El índice SRMR (en inglés Standardized Root Mean Square Residual) también es una medida global de ajuste del modelo, se puede interpretar como un error cuadrático medio, y se calcula cómo la raíz cuadrada del promedio de la suma de los residuos estandarizados. Los residuos son las diferencias entre la matriz de covarianzas muestral S , y la matriz de covarianzas reproducida por el modelo $\hat{\Sigma}_0$, p es el número de variables observadas.

$$SRMR = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^i (s_{ij} - \hat{\sigma}_{ij})^2 / (s_{ii}s_{jj})}{p(p+1)/2}}$$

Donde $(\hat{\sigma}_{ij} = \hat{\Sigma}_0)_{ij}$.

Índice de Buen Ajuste GFI

El índice GFI (Goodness of Fit Index), es un índice que mide cuanto mejora el modelo, en relación al modelo nulo que no tiene caminos (es decir el modelo en el cual todas las cargas y todas las covarianzas son 0, i.e. A y P son nulas). Al estadístico chi-cuadrado del modelo nulo lo denotamos χ_{nulo}^2 , con esto el índice se define como:

$$GFI = 1 - \frac{\chi^2}{\chi_{nulo}^2}$$

Este índice se puede interpretar como el porcentaje de varianza y covarianza de la matriz de covarianzas muestral que es predicho por la matriz de covarianzas reproducida por el modelo.

Criterio de Información de Akaike

Siempre al agregar parámetros a un modelo, este modelo mejora, sin embargo se pierde la parsimonia o simplicidad. Por esto el criterio de información de Akaike es un índice que considera el ajuste y el número de parámetros estimado.

El criterio de información de Akaike se define como:

$$AIC = \chi^2 + 2t$$

Donde t es el número de parámetros estimados del modelo. La idea es elegir el modelo con menor AIC . Esto maximiza el ajuste (dado que tiene el término de χ^2) pero también toma en consideración la complejidad del modelo ya que suma 2 por cada coeficiente nuevo estimado.

2.4. Caso de datos dicotómicos

En los análisis factoriales exploratorio y confirmatorio es importante el supuesto de normalidad de las variables observadas cuando los modelos se estiman con máxima verosimilitud, o al interpretar ciertos estadísticos de ajuste. Este supuesto en el caso de ítems dicotómicos no se cumple. Como ambos análisis usan la matriz de correlaciones para estimar los coeficientes del modelo, se plantea un enfoque que permite estimar correlaciones de variables que se distribuyen normales y utilizar posteriormente estas correlaciones en los análisis factoriales exploratorios o confirmatorios.

Se plantea un modelo alternativo, que es similar a los modelos logísticos, en donde se supone que existen variables x_i^* que definen a las variables originales x de la siguiente forma:

$$x_i = \begin{cases} 1 & \text{si } x_i^* > t_i, \\ 0 & \text{si } x_i^* \leq t_i. \end{cases} \quad (2.66)$$

Los parámetros t_i son umbrales, estos son desconocidos.

Bajo este enfoque, el objetivo es hacer los análisis, ya no sobre las variables x_i que son dicotómicas, si no que sobre las variables x_i^* que son continuas. Pero para aplicar métodos de análisis factorial con las variables x_i^* no es necesario conocer estas variables, si no que su matriz de correlaciones. Luego se estiman las correlaciones de las variables x_i^* a través de los datos observados de las variables x_i , estas correlaciones estimadas son las correlaciones tetracóricas de las variables x_i , las cuales se explicarán en detalle en el siguiente apartado.

2.4.1. Correlaciones Tetracóricas

La correlación tetracórica entre dos variables dicotómicas x_i y x_j (en este caso, dos ítems) es la estimación de la correlación entre las variables continuas x_i^* y x_j^* , que se supone que fueron dicotomizadas en las variables x_i y x_j de la siguiente forma:

$$x_i = \begin{cases} 1 & \text{si } x_i^* > t_i, \\ 0 & \text{si } x_i^* \leq t_i. \end{cases} \quad (2.67)$$

$$x_j = \begin{cases} 1 & \text{si } x_j^* > t_j, \\ 0 & \text{si } x_j^* \leq t_j. \end{cases} \quad (2.68)$$

Para estimar estas correlaciones se supone que la distribución conjunta de x_i^* y x_j^* es normal bivariada.

Consideremos la siguiente tabla de contingencia

x_i vs x_j	1	0
1	n_{11}	n_{10}
0	n_{01}	n_{00}

Con esto se tiene que la frecuencia relativa observada de que un par de realizaciones de las variables x_i y x_j caiga en el casillero kl con $k, l \in \{0, 1\}$ es $P_{kl} = n_{kl}/N$ donde N es el total

de datos. Se denota π_{kl} a las frecuencias relativas esperadas, dadas por la distribución de (x_i^*, x_j^*) y los umbrales t_i y t_j .

La idea es encontrar una distribución normal multivariada para las variables (x_i^*, x_j^*) y los umbrales t_i y t_j que definen a las variables dicotómicas. La distribución y los umbrales se calculan de forma que las frecuencias esperadas por celda π_{kl} sean lo más parecidas posibles a las frecuencias observadas P_{kl} . Se considera además el siguiente supuesto:

$$(x_i^*, x_j^*) \sim N\left(\hat{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \quad (2.69)$$

Como la media y la varianza de las distribuciones están fijas, el único parámetro que hay que estimar es ρ el cual es precisamente la correlación tetracórica entre x_i y x_j . Como (x_i^*, x_j^*) se distribuyen como una normal bivariada, su densidad es la siguiente:

$$\phi(x_i^*, x_j^*; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x_i^{*2} - 2\rho x_i^* x_j^* + x_j^{*2})\right] \quad (2.70)$$

Luego, la fórmula de la verosimilitud es la siguiente (la probabilidad de haber obtenido esas frecuencias observadas dado los umbrales y la correlación ρ):

$$L(n_{00}, n_{10}, n_{10}, n_{10}; t_i, t_j, \rho) = C \cdot \prod_{k=0}^1 \prod_{l=0}^1 \pi_{kl}^{n_{kl}} \quad (2.71)$$

Donde C es una constante, aquí podemos tomar el logaritmo (lo que no influye, ya que vamos a maximizar la verosimilitud, y el logaritmo es una función monótona creciente y no cambia el lugar donde se alcanza el máximo)

$$l = \ln(L) = \ln(C) \sum_1^{k=0} \sum_1^{l=0} n_{kl} \cdot \pi_{kl} = \ln(C) + n_{00}\pi_{00} + n_{10}\pi_{10} + n_{01}\pi_{01} + n_{11}\pi_{11} \quad (2.72)$$

La expresión exacta de los π_{kl} , considerando el supuesto de distribución normal bivariada de las variables dicotómicas, es:

$$\begin{aligned} \pi_{00} &= \int_{-\infty}^{t_i} \int_{-\infty}^{t_j} \phi(x, y; \rho) dx dy & \pi_{11} &= \int_{t_i}^{\infty} \int_{t_j}^{\infty} \phi(x, y; \rho) dx dy \\ \pi_{10} &= \int_{t_i}^{\infty} \int_{-\infty}^{t_j} \phi(x, y; \rho) dx dy & \pi_{01} &= \int_{-\infty}^{t_i} \int_{t_j}^{\infty} \phi(x, y; \rho) dx dy \end{aligned}$$

Luego para obtener la estimación de la correlación tetracórica ρ y los umbrales, que maximice la verosimilitud dado que se obtuvieron las frecuencias observadas n_{kl} con $k, l \in \{0, 1\}$ hay que resolver el siguiente problema de optimización:

$$\min_{t_i, t_j, 0 \leq \rho \leq 1} l(t_i, t_j, \rho) \quad (2.73)$$

La solución de este problema entrega los estimadores de máxima verosimilitud de ρ , t_i y t_j . Este cálculo es computacionalmente costoso y existen métodos menos precisos que se demoran menos. Sin embargo el software utilizado usa la estimación por máxima verosimilitud, por lo tanto no se ahondará en los otros métodos.

Los resultados y descripciones de esta sección son una adaptación de lo expuesto en [31] para correlaciones policóricas (que son similares a las tetracóricas, pero se aplican a variables con más categorías).

2.5. Resultados sobre las pruebas CAM y LMT

En esta sección se presenta el análisis factorial realizado sobre la prueba CAM para determinar las dimensiones que mide. A partir de este análisis se seleccionaron ítems para lograr una medida precisa del constructo. Se muestra también la relación de la medida dada por la prueba CAM con los resultados de la prueba de LMT sobre *conocimiento matemático común y específico*. Para esto se proyectaron los ítems de la prueba de LMT sobre el análisis factorial de la prueba CAM. La última parte de esta sección contiene los modelos de análisis factorial confirmatorio aplicados para testear la hipótesis de que estas pruebas permiten distinguir el constructo *conocimiento de alumnos y matemática* del constructo *conocimiento matemático común y específico*.

2.5.1. Análisis factorial de la prueba CAM

El objetivo de esta memoria es analizar la prueba CAM, entender su estructura y validarla. Por esto se realizaron análisis de componentes principales y factoriales sobre esta prueba para determinar que dimensiones o constructos está realmente midiendo. Para ver como se comporta la prueba de LMT en relación a la prueba CAM se proyectó la prueba LMT sobre el análisis factorial realizado sobre la prueba CAM. .

Los cálculos se hicieron utilizando diversos métodos de estimación, para demostrar que la conclusiones son estables y que no dependen del método utilizado.

Elección del número de factores

Se empezó el análisis de forma exploratoria realizando un ACP, ya que no depende del número de factores extraídos. Ahora bien, dado la gran cantidad de variables (61) no debieran de presentarse mayores diferencias entre los métodos de estimación.

Se realizaron dos análisis de componentes principales, uno con la matriz de correlaciones de Pearson y el otro con la matriz de correlaciones tetracóricas. Para determinar el número de factores a extraer se presentan: el gráfico de sedimentación y el resultado del análisis paralelo para ambas matrices de correlaciones (ver Figuras 2.1 y 2.2). Se puede observar que en la Figura 2.1, el *scree test* y el análisis paralelo coinciden en la conclusión de que la cantidad óptima de factores a extraer es 5. En la Figura 2.2 donde se utilizó la matriz de correlaciones tetracóricas, la cantidad óptima de factores sería 6. No se utilizó el criterio de Kayser y Guttman de considerar los valores propios mayores que 1, ya que como se observa en los gráficos de sedimentación esto produciría un extracción de demasiados factores.

Como se dijo anteriormente para comparar resultados y confirmar la estabilidad de las conclusiones, además del ACP, se realizó un análisis factorial usual con los tres métodos de estimación descritos en la sección 2.1: ejes principales (EP), máxima verosimilitud (MV) y residuos mínimos (MinRes).

El primer paso para realizar un análisis factorial es determinar el número m de factores comunes, ya que a diferencia del ACP, la estimación de las matrices Λ y L^2 depende de

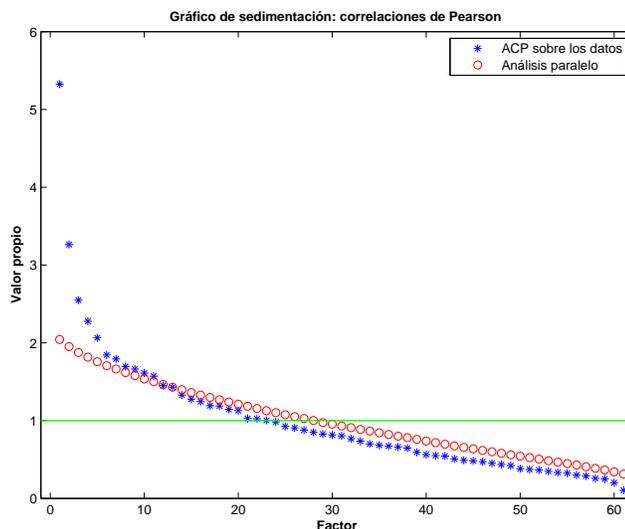


Figura 2.1: Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones de Pearson.

este número. Para determinar este número se utilizó la cantidad de factores sugerida por el análisis de componentes principales. Esto pues la varianza total es siempre mayor a la varianza común de las variables. Luego, si se puede representar bien la varianza total con un número m de factores, ese número debería ser suficiente para representar bien la varianza que es común. Considerando este razonamiento se realizó un análisis factorial para cada matriz de correlaciones, con la de Pearson se extrajeron 5 factores y con la de correlaciones tetracóricas se extrajeron 6. Para corroborar la decisión respecto al número de factores se usó el procedimiento de *scree plot* y un análisis paralelo específico para análisis factorial (el anterior era para análisis de componentes principales), y se confirmó la decisión sobre la cantidad de factores.

La comparación de los resultados obtenidos con los tres métodos de estimación se presenta en la siguiente subsección. Se espera teóricamente que la prueba sea unidimensional, es decir, que se explique bien con un sólo factor y que este coincida con el constructo teórico. Sin embargo, los resultados muestran que para explicarla bien es necesario considerar más de un factor.

Comparación de métodos de estimación

Se presenta la comparación de los resultados entregados por el análisis de componentes principales, máxima verosimilitud, ejes principales y residuos mínimos aplicados sobre la matriz de correlaciones de Pearson y tetracóricas.

Las principales conclusiones respecto a la estructura factorial de las variables se obtienen al observar las correlaciones de las variables con cada factor. Esta información se presenta en la matriz de estructura, que en el caso de factores incorrelados coincide con la de cargas. Esta matriz en su k -ésima columna tiene las correlaciones del factor k con cada variable observada. Por esto se identifica a cada columna con el factor del cual contiene a las correlaciones (a la

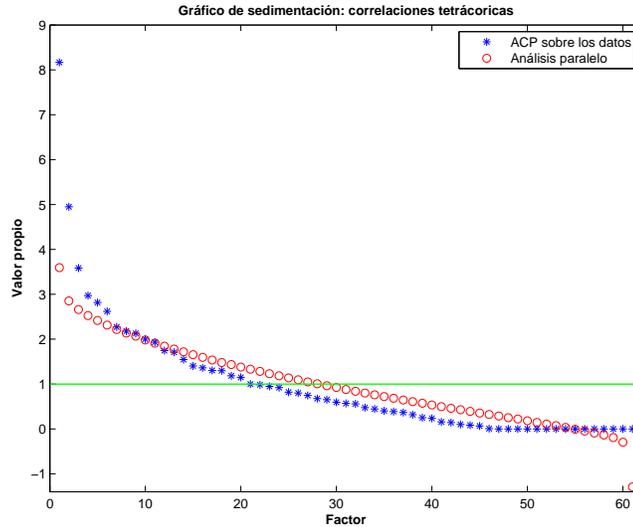


Figura 2.2: Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones tetracóricas.

primera columna con el factor 1 o f_1 , a la segunda con el factor 2 o f_2 , etc.).

Las columnas de la matriz de estructura permiten interpretar el significado de cada factor al observar con que variables el factor se correlaciona de manera alta o baja. Es por esto que para ver si los resultados llevan a conclusiones similares se compararon las columnas de las matrices de estructura entregadas por diversos métodos.

La comparación se hizo calculando la correlación entre la columna k de la matriz de estructura dada por un método con la misma columna de la matriz de estructura producida por otro método.

Los resultados de esta comparación para los análisis sobre la matriz de correlaciones de Pearson y tetracóricas se presentan respectivamente en las Tablas 2.1 y 2.2 que en cada fila indican los métodos comparados y en cada columna el factor representado por la columna de la matriz de estructura que se comparó. Los coeficientes que tienen son correlaciones, por ejemplo el primer coeficiente de la Tabla 2.1 representa la correlación entre la columna 1 de la matriz de estructura estimada con máxima verosimilitud con la columna 1 de la matriz de estructura estimada con ejes principales. Que esa correlación sea alta significa que el factor 1 en ambos análisis se interpreta de forma similar.

Métodos comparados	f_1	f_2	f_3	f_4	f_5
MV-EP	0,54	0,69	-0,53	0,81	0,97
MV-MINRES	1	1	1	1	1
EP-MINRES	0,55	0,69	-0,53	0,81	0,97
ACP-MV	0,56	0,62	-0,54	0,53	0,93
ACP-EP	0,99	0,99	0,92	0,9	0,99
ACP-MINRES	0,56	0,62	-0,55	0,53	0,93

Tabla 2.1: Correlaciones entre las columnas de las matrices de estructura entregadas por los métodos ACP, MV, EP y MinRes aplicados a la matriz de correlaciones de Pearson.

Métodos comparados	f_1	f_2	f_3	f_4	f_5	f_6
MV-EP	0,7	-0,77	-0,33	0,62	0,01	0,26
MV-MINRES	0,7	-0,76	-0,4	0,61	0,17	-0,01
EP-MINRES	1	1	0,99	0,96	0,93	-0,94
ACP-MV	0,7	-0,76	-0,4	0,61	0,17	-0,01
ACP-EP	1	1	0,99	0,96	0,93	-0,94
ACP-MV	1	1	1	1	1	1

Tabla 2.2: Correlaciones entre las columnas de las matrices de estructura entregadas por los métodos ACP, MV, EP y MinRes aplicados a la matriz de correlaciones tetracóricas.

De las Tablas se puede concluir que en el caso del análisis sobre la matriz de correlaciones de Pearson, los únicos métodos que podrían entregar resultados que lleven a conclusiones diferentes son ACP y ML (debido a que las correlaciones entre ambos son bajas, pero son altas con el resto). En el caso del análisis sobre la matriz de correlaciones tetracóricas se desprende la misma conclusión. Por lo tanto, se compararán las conclusiones a las que llevan estos dos métodos para ambas matrices.

Interpretación de factores

Para interpretar los factores, es necesario considerar que cada ítem se diseñó con el objetivo de que midiera al constructo *conocimiento de alumnos y matemáticas*. Por lo tanto, la validez de contenido de cada ítem ya se tiene por el procedimiento de diseño descrito en [45]. Sin embargo, al comparar los ítems con diversos análisis se distinguieron los que funcionaron mejor de los que no.

Otro punto previo importante a considerar, es el resultado entregado por Takane y De Leeuw² [40], que demuestra que los coeficientes de discriminación de IRT coinciden teóricamente con las cargas del primer factor de un modelo factorial sobre una matriz de correlaciones tetracóricas. Este resultado es teórico, sin embargo se reproduce bastante bien con los datos estimados. Luego, los ítems que mas carguen en el primer factor, son en general, los más discriminantes.

Primero se presenta la interpretación del análisis de componentes principales y se usan las conclusiones para interpretar el análisis factorial estimado con máxima verosimilitud.

Se presentan juntas las interpretaciones realizadas del ACP sobre la matriz de correlaciones de Pearson y sobre la matriz de correlaciones tetracóricas ya que sus resultados son prácticamente iguales (la correlación entre la primera columna de la matriz de estructura calculada con la matriz de correlaciones de Pearson con la columna análoga calculada con la matriz de correlaciones tetracóricas es 0,99). La única diferencia importante entre ambos resultados es que las cargas en el análisis con la matriz de correlaciones tetracóricas son más altas (ver Tablas A.2 y A.4). Este hecho es importante, ya que agrega ítems a la interpretación (ya que los ítems que explican un factor son los que tienen correlaciones altas con él).

²Este resultado se describe en la Sección 4.3

El primer factor tiene correlaciones positivas y altas en muchísimos más ítems que el resto de los factores. Además este factor tiene correlaciones positivas o cercanas a cero con todas las variables (no es bipolar). El resto de los factores representa principalmente grupos de ítems con comportamientos similares como por ejemplo los ítems 20a, 20b, 20c, 20d, o los ítems 13a, 13c, 13d.

Estas primeras observaciones sugieren que el primer factor estaría representando al constructo y el resto de los factores serían poco relevantes teóricamente. Para verificar que el primer factor representa al constructo *conocimiento de alumnos y matemáticas* se realizó un análisis detallado de los ítems que cargan de manera alta y baja en el factor:

Los ítems con correlaciones altas con el primer factor (30 primeros ítems de la primera columna de las Tablas A.2 y A.4) se distribuyen en las distintas áreas que componen al constructo representándolo bien. La descripción de estas áreas y los ítems que cargaron de manera alta al primer factor y las representan se describen en la siguiente lista[45]:

1. Ítems donde hay que identificar errores comunes y diseñar distractores: Por ejemplo los ítems 10b, 10e, 18a, 27. Los tres primeros buscan identificar si posibles distractores son adecuados o no, el último profundiza la tarea y pide razonar respecto a la causa que hace conveniente elegir un determinado distractor.
2. Ítems donde hay que reconocer causas de errores e identificar preconcepciones erróneas: Por ejemplo los ítems 19a, 19b, 19c, 21, 24c y 24d . Hay varios matices dentro de este tipo de preguntas, por ejemplo algunas piden identificar causas de error que son posibles, pero que al reflexionar sobre lo que produjo el estudiante son poco probables (19a, 19c, 21). Otros ítems preguntan por confusiones o preconcepciones comunes en los estudiantes, que un profesor con experiencia debiera de identificar como causantes de errores (19b, 24c, 24d). También existían ítems que pedían identificar los procedimientos o decisiones tomados por el profesor que facilitaron que el alumno cometiera un error (por ejemplo el ítem 21).
3. Identificar que ejercicios son más difíciles para un estudiante (entender como pensaría en cada uno de ellos): Por ejemplo el ítem 25 pide ordenar diversos polígonos según su dificultad en el cálculo de áreas.
4. Saber como convencer a un niño, que argumentos son suficientes y funcionan en su lógica: Por ejemplo los ítems 18c, 18d y 23, los ítems 18c y 18d preguntan si ciertos argumentos son suficientes para convencer a un niño. El ítem 23 pregunta por la mejor estrategia para que los niños comprendan que dos ángulos son iguales.
5. Conciencia acerca de ventajas, limitaciones o errores inducidos por representaciones o modelos usuales, así como ejemplos, definiciones o afirmaciones: Por ejemplo los ítems 16a, 17a, 17b, 17c, 26a, 26c, 26d, 28a, 28b, 28c. Esta categoría tiene varias subcategorías, por ejemplo los ítems 16a, 17a, 17b y 17c se refieren a las limitaciones y ventajas que puede generar el uso de ciertos modelos o metáforas. Los ítems 26a, 26c y 26d apuntan a identificar frases que pueden inducir errores y los ítems 28a, 28b, 28c y 28d piden identificar si distintas decisiones tomadas para enseñar una materia influyen en que los alumnos no sean capaces de aplicarla.

Respecto a los ítems con correlaciones cercanas a 0 en el primer factor, es posible plantear hipótesis respecto a características que hayan impedido que midan adecuadamente el

constructo. En el análisis con correlaciones de Pearson y tetracóricas, de los 10 ítems con menor correlación con el factor 1, 9 coincidieron en ambos análisis, las hipótesis que permiten entender por qué no se correlacionaron con el factor 1 son las siguientes:

1. Preguntas respecto a identificar si distractores son adecuados, a pesar que tres preguntas de distractores resultaron muy correlacionadas con el factor 1 (ítem 10b, 10e y 18a), aparecieron también tres preguntas similares (5, 10a y 18b que pedían identificar si un distractor era adecuado para un problema) que son prácticamente incorreladas con el primer factor, esto puede ser porque los distractores elegidos en las preguntas con alta correlación eran comunes y posibles de identificar para un profesor con alto *conocimiento de alumnos y matemática*. Probablemente las preguntas de distractores que resultaron poco correlacionadas tenían distractores no tan naturales, o que era menos claro si eran adecuados o no.
2. Respecto a las preguntas 1a, 1b, 1c y 9 una explicación posible para su poca correlación con el factor 1, es que la estrategia que había que analizar en las preguntas 1a-1b-1c no fuera común para los profesores, y que incluso para los mas expertos en *conocimiento de alumnos y matemática* fuera difícil identificar las dificultades de los niños que se abordaban con ella, ya que no eran muy usadas. Por otro lado la pregunta 9 pide hacer un proceso de reflexión muy profundo, que es el de distinguir que procedimientos que se mecanizan en el trabajo con una materia (en este caso proporciones) pueden influir en el desarrollo o manejo de otra materia relacionada (en este caso fracciones). Quizás esta tarea aún no es común para los profesores, incluso para los que tienen mucho *conocimiento de alumnos y matemática*. De hecho fue una pregunta difícil y sólo un 12.9% la respondió correctamente.
3. En el ítem 11 ocurre que a pesar de que la respuesta correcta es la más precisa, los otros distractores no son completamente falsos. Esto puede generar que la prueba mida “otros” conocimientos de alumnos y matemática [22], por ejemplo en el ítem 11 se quería que los profesores identificaran una causa de error común, que está reportada en la literatura, pero aparecieron otras causas factibles. Un comportamiento similar ocurre en el problema 4, donde los distractores tienen un componente de verdad.
4. La pregunta 13b pide identificar entre dos multiplicaciones cual es más difícil. Probablemente su baja correlación con el factor tiene que ver con que la respuesta correcta era muy sensible al orden en que el profesor enseña.

Se verifican los resultados teóricos de Takane y De Leeuw [40], la correlación de las cargas del primer factor con la discriminaciones de los ítems calculadas con teoría clásica fue 0.98 y con las discriminaciones calculadas por IRT fue 0,96. Por el comportamiento de los ítems más correlacionados con el factor, y porque se puede dar una explicación respecto a los que correlacionan poco con éste: El primer factor en ambos análisis se interpreta como el que representa al constructo *conocimiento de alumnos y matemática*. Este resultado es muy importante, ya que comprueba que a pesar de la presencia de factores sin una interpretación teórica sustantiva, se pudo identificar al constructo con la prueba.

Las interpretaciones anteriores fueron realizadas sobre el análisis de componentes principales, se presenta ahora la interpretación de los resultados entregados por el método de estimación de máxima verosimilitud. El primer factor entregado por el método de máxima verosimilitud en ambas matrices de correlaciones se presenta como una mezcla entre el factor

1 y el factor 3 de ACP. Por lo tanto se estarían mezclando factores que en el ACP se presentaban separados, en particular por la interpretación del ACP, uno de los factores mezclados es el que representa al constructo *conocimiento de alumnos y matemática*. Por eso se aplicó una rotación Promax a ambos análisis.

Después de la rotación se obtuvo que el primer factor (el con mayor varianza) tenía correlaciones altas con los mismos ítems que el primer factor del análisis de componentes principales, factor que como se vió, representaba al constructo *conocimiento de alumnos y matemática*. Esto ocurrió con ambas matrices de correlaciones. El resto de los factores entregados por máxima verosimilitud con rotación Promax, también representan conjuntos de ítems que se comportan de manera similar. En conclusión usando como método de estimación máxima verosimilitud se identificó al factor que coincidía con el constructo pero después de una rotación.

Selección de ítems

Se van a seleccionar a los ítems con mayores cargas o mayores correlaciones con el factor que representa al *conocimiento de alumnos y matemática*. Esto se hará considerando que aparezcan formando este factor de manera robusta (es decir con todos los métodos de estimación y con las dos matrices de correlaciones).

Esto también entrega los ítems con mayor discriminación por la alta correlación de la primera columna de la matriz de estructura con las discriminaciones. Luego la selección resultante de este proceso debiera de presentar excelentes propiedades psicométricas. Se consideraron los 30 ítems con correlaciones más altas en el factor 1 en los 4 análisis donde se encontró que el factor 1 representaba al constructo. Se hizo la selección sobre los primeros 30 ítems ya que aproximadamente (como se ve en la Tabla 2.3) tienen correlaciones con el factor 1 mayores a 0,3 .

Método	Matriz de Correlaciones	Valor mínimo
ACP	Tetracóricas	0,35
ACP	Pearson	0,29
MV - Promax	Tetracóricas	0,27
MV - Promax	Pearson	0,21

Tabla 2.3: Para cada modelo factorial estimado se tomaron los 30 ítems que correlacionaron más alto con el factor 1. En la tabla se presenta el mínimo de estas correlaciones en cada modelo (los modelos se identifican con el método y la matriz usada en la estimación).

Es importante notar que si no se hubiera realizado el análisis con la matriz de correlaciones tetracóricas se habrían seleccionado menos ítems ya que con la matriz de correlaciones de Pearson las correlaciones entre los factores y los ítems fueron menores.

Se seleccionaron los ítems que estaban en el grupo de los 30 primeros en los 4 análisis simultáneamente, estos son: I10b, I10e, I16a, I17b, I18a, I18c, I18d, I19a, I19b, I19c, I21, I23, I24a, I24b, I24c, I25, I26a, I26c, I26d, I27, I28a, I28b, I28c, I28e. Sin embargo se agregaron además los siguientes ítems: I24d y 17d que estaban entre los 30 primeros en 3 de los 4 análisis.

Por último se agregaron los ítems I17a y I15 que también aparecieron entre los primeros 30 en más de un análisis.

Análisis factorial de los ítems seleccionados

Se hicieron análisis de componentes principales de los 28 ítems seleccionados, con la matriz de correlaciones de Pearson y tetracóricas. Los gráficos de sedimentación para cada matriz de correlaciones se presentan en las Figuras 2.3 y 2.4.

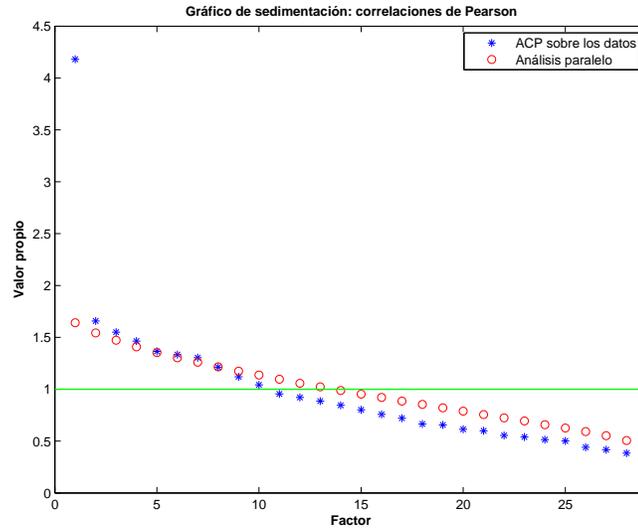


Figura 2.3: Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones de Pearson con la selección de ítems.

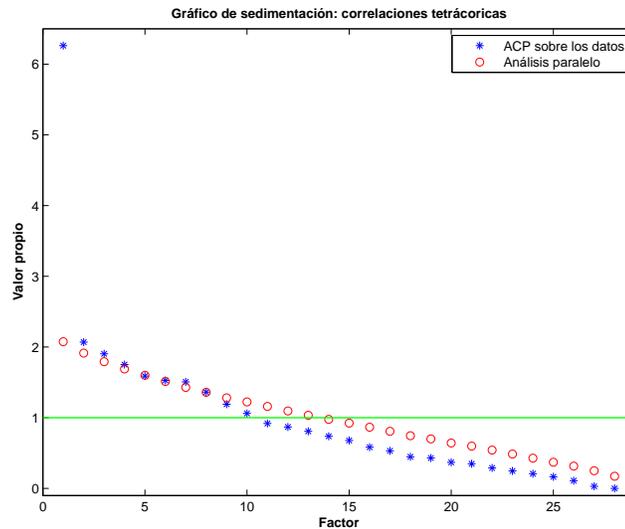


Figura 2.4: Gráfico de sedimentación del análisis realizado sobre la matriz de correlaciones tetracóricas con la selección de ítems.

En los gráficos de sedimentación (Figuras 2.3 y 2.4) se puede observar que la prueba depurada es unidimensional para ambas matrices. Los coeficientes del primer factor para

cada matriz se presentan en las Tablas 2.4 y 2.5.

Ítem	Correlación con f_1						
I10b	0,38	I18a	0,31	I23	0,32	I26c	0,47
I10e	0,34	I18c	0,3	I24a	0,47	I26d	0,33
I15	0,22	I18d	0,29	I24b	0,39	I27	0,52
I16a	0,38	I19a	0,45	I24c	0,32	I28a	0,29
I17a	0,23	I19b	0,48	I24d	0,29	I28b	0,42
I17b	0,25	I19c	0,44	I25	0,48	I28c	0,45
I17d	0,27	I21	0,39	I26a	0,47	I28e	0,58

Tabla 2.4: Correlación de cada ítem con el factor 1 calculado con un ACP sobre la matriz de correlaciones de Pearson.

Ítem	Correlación con f_1						
I10b	0,45	I18a	0,37	I23	0,39	I26c	0,56
I10e	0,41	I18c	0,38	I24a	0,58	I26d	0,4
I15	0,28	I18d	0,38	I24b	0,47	I27	0,65
I16a	0,49	I19a	0,54	I24c	0,38	I28a	0,39
I17a	0,27	I19b	0,6	I24d	0,35	I28b	0,51
I17b	0,31	I19c	0,54	I25	0,61	I28c	0,53
I17d	0,36	I21	0,46	I26a	0,55	I28e	0,68

Tabla 2.5: Correlación de cada ítem con el factor 1 calculado con un ACP sobre la matriz de correlaciones tetracóricas.

Por lo mostrado en la interpretación del análisis de componentes principales sobre la prueba completa, esta selección de ítems representa ampliamente las diversas vetas del *conocimiento de alumnos y matemáticas*. Además, todos los ítems se correlacionan significativa y positivamente con el primer factor, el cual es el más relevante por el alto porcentaje de varianza que explica. Luego esta selección cumple con las exigencias técnicas (es unidimensional) y teóricas (representa al constructo).

Conclusiones

Se logró identificar al factor que representa al constructo *conocimiento de alumnos y matemática*, de manera robusta con todos los métodos utilizados y con ambas matrices. Sin embargo, aparecieron factores sin una interpretación teórica sustantiva que representaban principalmente similitud entre ítems. Esto motiva a depurar el instrumento, eliminando los ítems que no se correlacionan con el factor que representa al constructo y que generan factores no significativos teóricamente.

Se logró depurar el instrumento formando una subescala con buenas propiedades técnicas y con validez de contenido y medición. Esto permite usar esta medición para análisis que la comparen con otras mediciones y lograr determinar si se presenta validez de convergencia o divergencia con otros constructos.

Proyección de la prueba LMT sobre el análisis de la prueba CAM

Se proyectaron los ítems de la prueba de LMT sobre el ACP realizado sobre la selección de ítems. Se utilizó el ACP realizado sobre la matriz de correlaciones de Pearson, ya que en el análisis sobre correlaciones tetracóricas no se puede calcular el valor de cada factor en cada persona.

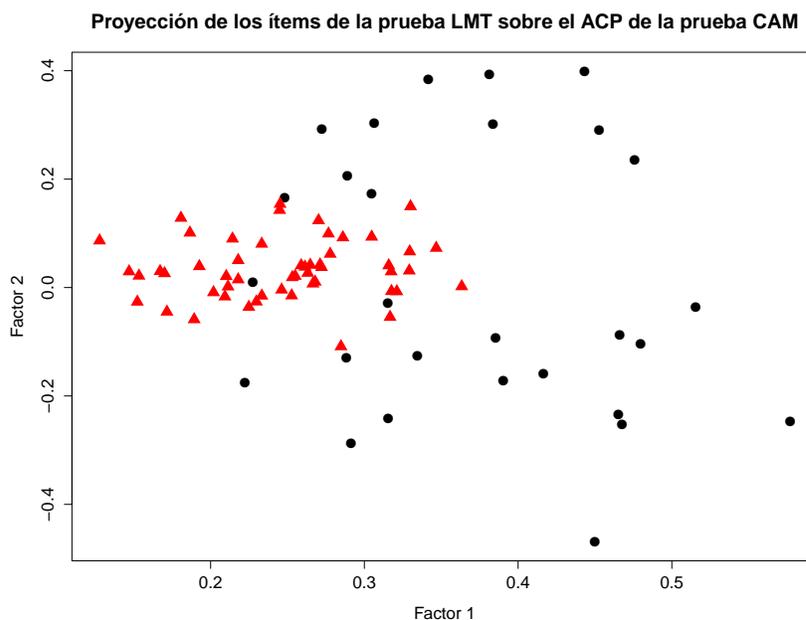


Figura 2.5: Proyección de los ítems de la prueba LMT sobre el ACP de la prueba CAM.

En la Figura 2.5 el eje horizontal representa las correlaciones de los ítems de las pruebas con el factor 1 del ACP sobre la prueba CAM depurada, y el eje vertical a las correlaciones de los ítems con el factor 2 del ACP sobre la prueba CAM depurada. Se presentan en triángulos rojos las correlaciones de los ítems que pertenecen a la prueba de LMT y en círculos negros los que pertenecen a la prueba CAM.

Si ambas pruebas midieran el mismo constructo, debiera de observarse una mezcla entre sus ítems. Si embargo esto no ocurre y se produce una separación clara, donde los ítems LMT tienen correlaciones bajas y positivas con el factor 1. Lo que tiene sentido, ya que ambas pruebas tienen en común un componente matemático.

2.5.2. Análisis factorial confirmatorio sobre la prueba CAM y la prueba LMT

Dentro de la teoría del *conocimiento matemático para enseñar* se espera que los constructos *conocimiento de alumnos y matemáticas* y *conocimiento matemático común y específico* sean distinguibles entre sí, pero que tengan correlación no nula. Este último supuesto es porque ambos constructos tienen una componente de habilidad y conocimiento matemático común. Es decir para tener un alto nivel de *conocimiento de alumnos y matemática* o de

conocimiento matemático común y específico, es necesario un mínimo de conocimientos y habilidades matemáticas.

Se van a evaluar las siguientes hipótesis:

- Los constructos son distinguibles y correlacionados entre sí, es decir, cada prueba forma un factor que representa al constructo correspondiente y los factores están correlacionados (ver Modelo 1 en Figura 2.6).
- Los constructos no son distinguibles (por lo menos con estos instrumentos), es decir ambas pruebas forman un factor común. (ver Modelo 2 en Figura 2.6)

En los modelos de análisis factorial confirmatorio es importante el supuesto de normalidad para interpretar adecuadamente los índices de ajuste. Por esta razón, en vez de trabajar con los ítems directamente, se formaron parcelas con cada prueba. Para formar las parcelas que representan cada instrumento se realizó una partición de los ítems de la prueba en k conjuntos. Para cada conjunto de la partición se define una parcela como la variable formada por la suma de los ítems del conjunto. Luego se resumen los p ítems en k parcelas. Como todos los ítems que forman a la prueba CAM representan significativamente al constructo medido (ya que todos tienen una alta correlación con el primer factor del análisis factorial) cualquier suma aleatoria de ítems representa bien al constructo medido.

Considerando esto se formaron 4 parcelas aleatorias con los ítems de la prueba CAM denotadas por PC1, PC2, PC3 y PC4. Cada una formada por la suma de 7 ítems. Para la prueba de LMT se formaron 7 parcelas denotadas PL1, PL2, PL3, PL4, PL5, PL6 y PL7. Las primeras 4 formadas por 7 ítems cada una, y las tres restantes formadas por 8 ítems cada una. A través de histogramas se verificó que las parcelas formadas en general se acercaban bastante al supuesto de normalidad. Con estas parcelas se testearon los modelos de análisis factorial confirmatorio presentados en la Figura 2.6.

Para comprobar que los resultados no dependieran de la elección aleatoria de los ítems de las parcelas, el proceso de formación aleatoria de parcelas y cálculo de los modelos 1 y 2 se realizó 100 veces. La comparación de los resultados se presenta en la Tabla 2.6.

Modelo 1		Modelo 2	
Índice	Promedio sobre los 100 intentos	Índice	Promedio sobre los 100 intentos
RMSEA	0,045*	RMSEA	0,119
SRMR	0,034*	SRMR	0,081**
GFI	0,956*	GFI	0,853
AIC	116,9	AIC	262.3

Tabla 2.6: Promedio de los índices de ajuste en las 100 generaciones aleatorias de parcelas para cada modelo.

Se marcó con un asterisco (*) los promedios de índices que representaban un ajuste bueno, y con dos asteriscos (**) los promedios de los índices que representan un ajuste aceptable según [35].

Para comparar de forma mas exigente el resultado de los modelos, se graficó la distribución de cada índice en cada modelo en los 100 cálculos de parcelas aleatorias. Los gráficos con la distribución de cada índice para cada modelo se presentan en la Figura 2.7. Se marcó con una

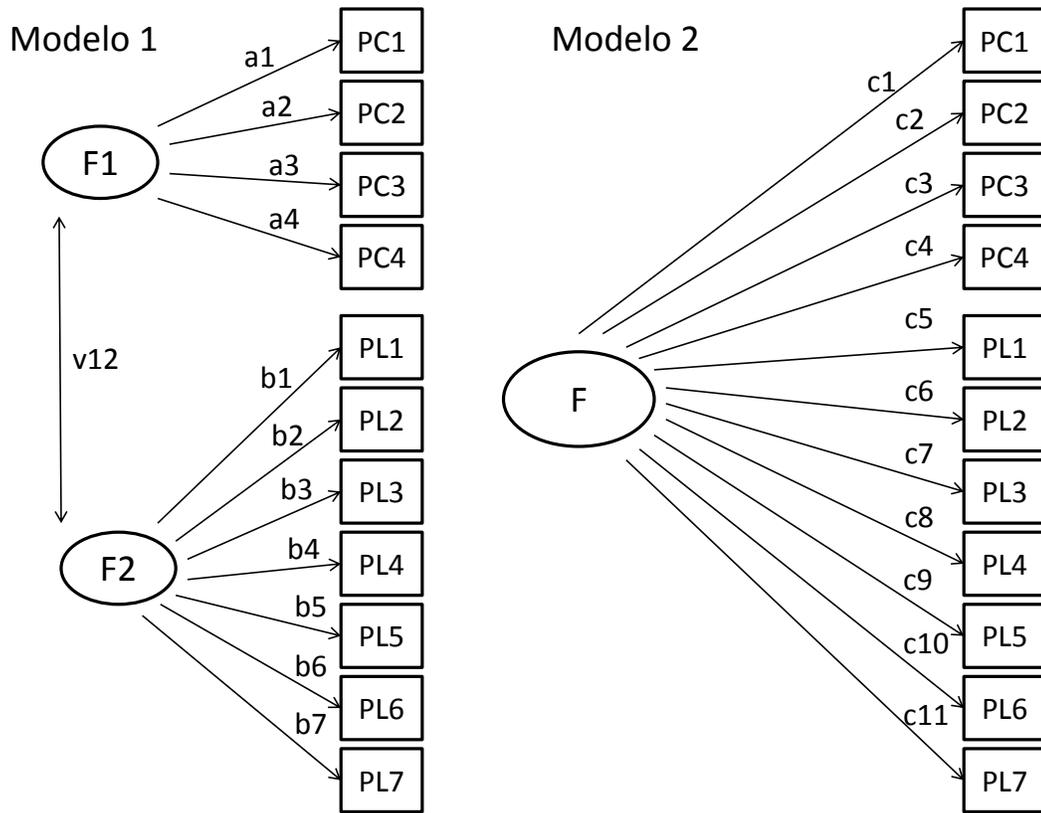
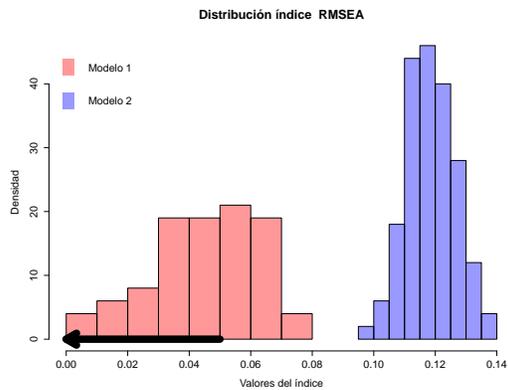


Figura 2.6: Modelos de análisis factorial confirmatorio.

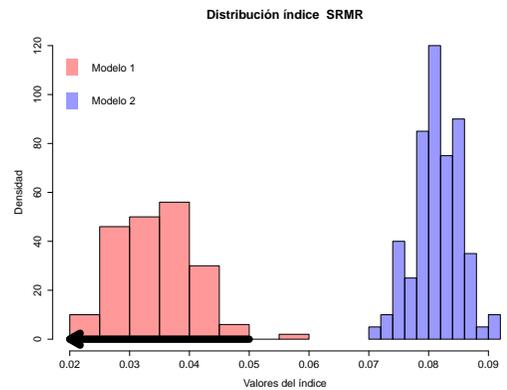
flecha los rangos de valores en los índices que muestran un índice de ajuste excelente [35].

En el Modelo 1 para las 100 estimaciones aleatorias de las parcelas los coeficientes a1, a2, a3, a4, b1, b2, b3, b4, b5, b6, b7, v12 fueron todos significativamente positivos (el menor valor que se alcanzó fue 0,52). En el Modelo 2 para las 100 estimaciones de las parcelas los coeficientes c1, c2, c3, c4, c5, c6, c7, c8, c9, c10, c11 fueron todos significativamente positivos (el menor valor alcanzado fue 0,35). Es importante que todas las parcelas tengan coeficientes positivos, ya que se supone que todas aportan a los respectivos factores, en ambos modelos.

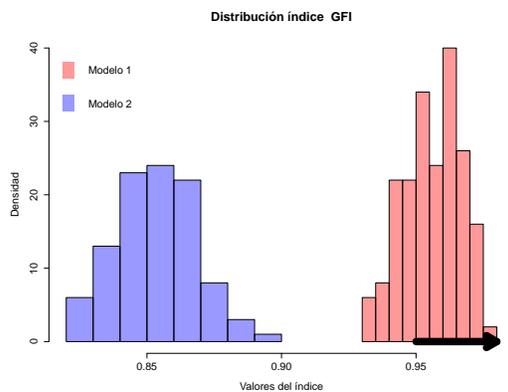
Los dos modelos representan a los datos como se espera (todas las parcelas cargan de forma positiva y significativa en los factores de ambos modelos), sin embargo el ajuste del Modelo 1 muestra que este representa de manera excelente a los datos, y además de forma muy superior al Modelo 2. Es más sólo el Modelo 1 alcanzó varias veces valores excelentes para los índices RMSEA, SRMR y GFI, y todos los modelos simulados (hasta el peor modelo) alcanzaron valores para el AIC mejores que en el Modelo 2. Por lo tanto se concluye que las mediciones proporcionadas por la prueba CAM y LMT permiten distinguir los constructos.



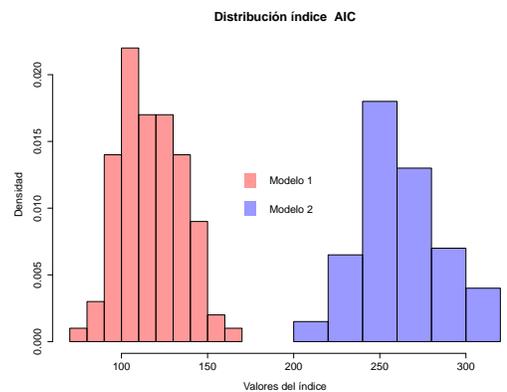
(a)



(b)



(c)



(d)

Figura 2.7: Distribución de los índices de ajuste en el Modelo 1 y el Modelo 2, para las 100 parcelas aleatorias que se estimaron.

Capítulo 3

Teoría Clásica de Test

La teoría clásica de test fue el primer marco teórico que permitió modelar puntajes de pruebas y analizar su error de medición y las propiedades de sus ítems. Aún se utiliza mucho, especialmente en lo que concierne a la confiabilidad de los test. El modelo se basa en los conceptos de puntaje verdadero y puntaje observado, los cuales son relacionados. En particular Lord and Novick dijeron [27]: “El modelo de TCT está basado en una definición particular, matemáticamente conveniente de puntaje verdadero y algunos supuestos básicos sobre las relaciones entre puntajes verdaderos y puntajes de error”. A partir de esa definición y de los supuestos del modelo se definen los conceptos de confiabilidad y error estándar de medición. En este capítulo se describen estos conceptos y como estimarlos en pruebas formadas por ítems. Se muestran dos enfoques para estimar confiabilidad, el tradicional de teoría clásica y uno más moderno que utiliza modelos de análisis factorial. Se explican también los parámetros que permiten evaluar los ítems, que son la dificultad y discriminación. Sólo se presentan los conceptos y resultados necesarios para analizar la prueba CAM. Sin embargo existen muchos más tópicos que la teoría clásica abarca como la validez y el estudio de pruebas que no están formadas por ítems (por ejemplo de velocidad).

3.1. Teoría Clásica en test formados por ítems

3.1.1. Descripción del modelo

Supongamos que se tiene un test¹ para medir un atributo. Este test le asigna a cada persona un puntaje. Llamaremos X a la variable aleatoria que representa el puntaje entregado por el instrumento, que cambia en cada persona.

Se supone que este puntaje está formado por dos componentes independientes: una variable latente T , que representa el puntaje real, y una variable de error E que representa el error de medición. Las variables T y E no son observables directamente. La ecuación principal de

¹Este test puede ser una prueba, un ítem, una encuesta, etc.

teoría clásica es:

$$X = T + E \quad (3.1)$$

Se agregan además los siguientes supuestos sobre T y E [15], donde se supone que se tienen dos test con puntajes observados $X_1 = T_1 + E_1$ y $X_2 = T_1 + E_2$:

$$\mathbb{E}(E) = 0 \quad (3.2)$$

$$\text{cor}(T, E) = 0 \quad (3.3)$$

$$\text{cor}(E_1, E_2) = 0 \quad (3.4)$$

$$\text{cor}(E_1, T_2) = 0 \quad (3.5)$$

Estos supuestos se basan en que el error es completamente aleatorio y que cualquier error sistemático fue eliminado como consecuencia del protocolo de medición [15]. A partir de estos supuestos se tiene que la esperanza del puntaje verdadero es la misma que la esperanza del puntaje observado:

$$\mathbb{E}(X) = \mathbb{E}(T) + \mathbb{E}(E) = \mathbb{E}(T) \quad (3.6)$$

Además se tiene que la varianza del puntaje observado es la suma de la varianza del puntaje verdadero y la varianza del error.

$$\text{var}(X) = \text{var}(T) + \text{var}(E) \quad (3.7)$$

La ecuación (3.7) es muy importante porque permite caracterizar el concepto de confiabilidad que se define como [27]: la fracción de la varianza total producida por el puntaje verdadero. Este concepto es uno de los más importantes en medición, pues permite tener una noción del efecto del error en los puntajes que se obtienen de un instrumento. Su definición matemática es la siguiente:

Definición 3.1 (Confiabilidad)

$$R_X = \frac{\text{var}(T)}{\text{var}(X)} = \frac{\text{var}(T)}{\text{var}(T) + \text{var}(E)} \quad (3.8)$$

Existen formas equivalentes de expresar este coeficiente que son fáciles de demostrar usando los supuestos:

$$R_X = \text{cor}(X, T) \quad (3.9)$$

$$R_X = 1 - \frac{\text{var}(E)}{\text{var}(X)} \quad (3.10)$$

Usando la definición de confiabilidad y las ecuaciones (3.9) y (3.10) se tiene que

1. La confiabilidad es un coeficiente que se mueve en un rango de 0 a 1
2. Es la fracción de varianza total que es producida por cambios en el puntaje verdadero
3. Si la confiabilidad es alta significa que las diferencias de puntaje observado entre dos individuos se deben principalmente a una diferencia entre el puntaje verdadero y no a una diferencia en el puntaje de error.
4. También, dado que es igual a la correlación entre X y T se tiene que la confiabilidad es una medida de la relación lineal entre estas variables.

5. Por la ecuación (3.10) se tiene que la confiabilidad también es una medida de cuan grande es el error en comparación al puntaje observado.

Con las ecuaciones expuestas no es evidente como se podría estimar el coeficiente de confiabilidad, ya que la única variable que se puede observar directamente es X . Por esto se definen los siguientes conceptos que relacionan a dos test y que permiten determinar distintos estimadores de confiabilidad [15]:

Definición 3.2 (Test paralelos) *Se dice que X_1 y X_2 son test paralelos, si se cumple que:*

$$X_1 = T + E_1 \quad (3.11)$$

$$X_2 = T + E_2 \quad (3.12)$$

$$\text{var}(E_1) = \text{var}(E_2) \quad (3.13)$$

Se puede inferir inmediatamente de la definición, que si dos test son paralelos:

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) \quad (3.14)$$

$$\text{var}(X_1) = \text{var}(X_2) \quad (3.15)$$

$$\text{cor}(X_1, X_2) = \frac{\text{cov}(T, T) + \text{cov}(T, E_1) + \text{cov}(T, E_2) + \text{cov}(E_1, E_2)}{\text{var}(X_1)} = \frac{\text{var}(T)}{\text{var}(X_1)} \quad (3.16)$$

Por la ecuación (3.16) se tiene que la confiabilidad de test paralelos es igual a la correlación entre sus puntajes observados:

$$R_{X_1} = R_{X_2} = \text{cor}(X_1, X_2) \quad (3.17)$$

La ecuación (3.17) es muy importante porque permite expresar la confiabilidad como una correlación de variables que son observables, y por lo tanto da una base para poder estimarla. Sin embargo el supuesto de que los test son paralelos es demasiado fuerte, por eso se han definido otras relaciones entre test menos restrictivas:

Definición 3.3 (Test tau-equivalentes) *Se dice que X_1 y X_2 son test tau-equivalentes, si se cumple que:*

$$X_1 = T + E_1 \quad (3.18)$$

$$X_2 = T + E_2 \quad (3.19)$$

$$(3.20)$$

Definición 3.4 (Test congénéricos) *Se dice que X_1 y X_2 son test congénéricos, si se cumple que:*

$$X_1 = \gamma_1 + \beta_1 T + E_1 \quad (3.21)$$

$$X_2 = \gamma_2 + \beta_2 T + E_2 \quad (3.22)$$

$$(3.23)$$

3.1.2. Confiabilidad en pruebas formadas por ítems

Consideremos ahora el caso particular, en que el puntaje observado es la suma de los puntajes de p ítems, es decir:

$$X = \sum_{i=1}^p X_i \quad (3.24)$$

En este caso existen muchos coeficientes que bajo ciertos supuestos son iguales a la confiabilidad del test X . Para definir estos coeficientes se supone que los ítems son congenéricos, que es el caso más general:

$$X_i = \gamma_i + \beta_i T + E_i \quad (3.25)$$

$$X = \sum_{i=1}^p (\gamma_i + \beta_i T + E_i) \quad (3.26)$$

Así pues, usando los supuestos descritos en (3.2), tenemos que la varianza del puntaje observado es:

$$\text{var}(X) = \underbrace{\sum_{i=1}^p \sum_{j=1}^p \beta_i \beta_j \text{var}(T)}_{\text{Varianza puntaje verdadero}} + \underbrace{\sum_{i=1}^p \text{var}(E_i)}_{\text{Varianza de error}} \quad (3.27)$$

De donde se puede distinguir una componente de varianza producida por el puntaje verdadero y otra por el error. Con esto la confiabilidad del puntaje total X sería:

$$R_X = \frac{\sum_{i=1}^p \sum_{j=1}^p \beta_i \beta_j \text{var}(T)}{\sum_{i=1}^p \sum_{j=1}^p \beta_i \beta_j \text{var}(T) + \sum_{i=1}^p \text{var}(E_i)} \quad (3.28)$$

Si llamamos Σ_X a la matriz de covarianzas entre los ítems, cuya dimensión es $p \times p$, esta tiene la siguiente expresión:

$$\Sigma_X = \begin{pmatrix} \beta_1^2 \text{var}(T) + \text{var}(E_1) & \beta_1 \beta_2 \text{var}(T) & \dots \\ \beta_2 \beta_1 \text{var}(T) & \beta_2^2 \text{var}(T) + \text{var}(E_2) & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (3.29)$$

Los supuestos de teoría clásica y el hecho de que los ítems sean congenéricos genera que la matriz de covarianzas de los ítems Σ_X sólo tenga varianza de error en su diagonal y los términos de afuera de la diagonal sólo representan varianza producida por el puntaje verdadero. Esto también permite intuir porque a medida que aumenta la cantidad de ítems, en general la confiabilidad aumenta también (si es que los ítems que se agregan correlacionan de manera positiva con el resto), ya que los elementos de afuera de la diagonal crecen en un orden p^2 y los de la diagonal crecen en un orden p .

Vamos a describir distintos coeficientes que se han inventado para estimar la confiabilidad a partir de pruebas formadas por ítems. Estos coeficientes en general son sólo cotas inferiores, y se transforman en la confiabilidad bajo ciertos supuestos. Los coeficientes que se van a describir en esta sección son el de split-half, y las cotas inferiores dadas por Guttman en 1945 [16] más famosas. Entre esas cotas ya estaba el coeficiente alfa de Cronbach (con el nombre de λ_3), el cual se popularizó después de que Cronbach publicara un paper en 1951

[9] en donde entregó resultados que permitían interpretarlo mejor. También se presentan el coeficiente ω_h propuesto por McDonald [28] que pide relaciones menos restrictivas entre los ítems para coincidir con la confiabilidad. Vamos a expresar de dos formas equivalentes a la confiabilidad para hacer evidente la relación entre esta y los coeficientes:

$$R_X = \frac{\sum_{i \neq j} \beta_i \beta_j \text{var}(T) + \sum_{k=1}^p \beta_k^2 \text{var}(T)}{\sum_{i=1}^p \sum_{j=1}^p \beta_i \beta_j \text{var}(T) + \sum_{i=1}^p \text{var}(E_i)} \quad (3.30)$$

$$R_X = \frac{\text{var}(X) - \sum_{k=1}^p \text{var}(E_k)}{\text{var}(X)} \quad (3.31)$$

1. Split-half: El coeficiente de split-half se basa en la idea de dividir la prueba en dos mitades paralelas, y estimar la confiabilidad como la correlación entre los puntajes producidos por estas dos mitades. Sin embargo esta correlación sería la confiabilidad de las mitades por separado, luego hay que corregirla para que represente la confiabilidad de la prueba total. Esta corrección se basa en el siguiente teorema y su corolario:

Teorema 3.5 *Consideremos Y_1 e Y_2 dos puntajes observados, y sea $X = Y_1 + Y_2$, entonces se tiene que:*

$$R_X \geq 2 \left[1 - \frac{\text{var}(Y_1) + \text{var}(Y_2)}{\text{var}(X)} \right] \quad (3.32)$$

DEMOSTRACIÓN. Véase [27] pp. 87-88. □

Corolario 3.6 *Si Y_1 e Y_2 son medidas paralelas, entonces:*

$$R_X = 2 \left[1 - \frac{\text{var}(Y_1) + \text{var}(Y_2)}{\text{var}(X)} \right] = \frac{2\text{cor}(Y_1, Y_2)}{1 + \text{cor}(Y_1, Y_2)} \quad (3.33)$$

A partir del corolario, se puede calcular la confiabilidad de la prueba aplicando la fórmula (3.33) a la correlación entre dos mitades. Sin embargo para que este término sea la confiabilidad y no sólo una cota inferior estas mitades deben ser paralelas.

2. λ_1 de Guttman: Este coeficiente, que es sólo una cota inferior de la confiabilidad, se define como:

$$\lambda_1 = 1 - \frac{\sum_{i=1}^p \text{var}(Y_i)}{\text{var}(X)} \quad (3.34)$$

Es una cota inferior, porque es equivalente a remplazar en la ecuación de la confiabilidad (3.30) a la componente $\sum_{k=1}^p \beta_k^2 \text{var}(T)$ por 0. Luego se tiene que:

$$R_X > \frac{\sum_{i \neq j} \beta_i \beta_j \text{var}(T)}{\sum_{i=1}^p \sum_{j=1}^p \beta_i \beta_j \text{var}(T) + \sum_{i=1}^p \text{var}(E_i)} = \lambda_1 \quad (3.35)$$

3. λ_2 de Guttman: Este coeficiente fue presentado por Guttman en [16]. Es muy interesante ya que es una cota inferior para la confiabilidad igual o mayor que el coeficiente alfa de Cronbach. Sin embargo es mucho menos famoso. Su formulación es la siguiente:

$$\lambda_2 = 1 - \frac{\sum_{i=1}^p \text{var}(X_i)}{\text{var}(X)} + \frac{\sqrt{\frac{p}{p-1} \sum_{i \neq j} \text{cov}(X_i, X_j)^2}}{\text{var}(X)} \quad (3.36)$$

En este caso, se puede ver que el término de la confiabilidad $\sum_{k=1}^p \beta_k^2 \text{var}(T)$ (ver ecuación (3.30)) se reemplaza por $\sqrt{\frac{p}{p-1} \sum_{i \neq j} \text{cov}(X_i, X_j)}$. El teorema que relaciona a λ_2 con la confiabilidad y el alfa de Cronbach es el siguiente:

Teorema 3.7 Sean X_1, X_2, \dots, X_p , p ítems que forman al puntaje X de una prueba, de forma que $X = \sum_{i=1}^p X_i$. Entonces

$$\lambda_2 \leq R_X \quad (3.37)$$

$$\alpha \leq \lambda_2 \quad (3.38)$$

DEMOSTRACIÓN. Para la demostración de la ecuación (3.37) véase [27] pp. 94-95.

Para la ecuación (3.38) véase [42]. □

4. α de Cronbach o λ_3 de Guttman: Este coeficiente fue propuesto por Guttman con la notación λ_3 en [16] y luego por Cronbach que lo llamó α en [9]. Este coeficiente es el más usado porque en general subestima la confiabilidad, es decir es conservador y en el caso que los ítems sean tau-equivalentes coincide con la confiabilidad. Su expresión es:

$$\alpha \text{ o } \lambda_3 = 1 - \frac{\sum_{i=1}^p \text{var}(Y_i)}{\text{var}(X)} + \frac{\text{var}(X) - \sum_{i=1}^p \text{var}(X_i)}{(p-1)\text{var}(X)} \quad (3.39)$$

La expresión (3.39), aunque parece un poco complicada, permite hacer evidente, que el término $\sum_{k=1}^p \beta_k^2 \text{var}(T)$ de la confiabilidad en este caso se reemplaza por $(\text{var}(X) - \sum_{i=1}^p \text{var}(X_i))/(p-1)$.

Se tienen los siguientes teoremas que relacionan a este coeficiente con la confiabilidad, el primero es directo a partir del Teorema 3.7, sin embargo fue primero que este último y es muy importante, por esto se presenta aparte.

Teorema 3.8 Sean X_1, X_2, \dots, X_p , p ítems que forman al puntaje X de una prueba, de forma que $X = \sum_{i=1}^p X_i$. Entonces

$$R_X \geq \alpha \quad (3.40)$$

DEMOSTRACIÓN. Véase [27] pp. 88-90. □

Teorema 3.9 Si los ítems X_1, X_2, \dots, X_p son medidas tau-equivalentes entre sí, entonces:

$$R_X = \alpha \quad (3.41)$$

DEMOSTRACIÓN. Para demostrar que α coincide con la confiabilidad basta demostrar que

$$\sum_{i=1}^p \beta_i^2 \text{var}(T) = \text{var}(X) - \sum_{i=1}^p \text{var}(X_i) = \sum_{i=1}^p \frac{1}{p-1} \sum_{j=1, j \neq i}^p \text{cov}(X_i, X_j)$$

Por un lado, como los ítems son tau-equivalentes

$$\sum_{i=1}^p \beta_i^2 \text{var}(T) = \sum_{i=1}^p \text{var}(T)$$

Por la misma razón

$$\text{var}(X) - \sum_{i=1}^p \text{var}(X_i) = \sum_{i=1}^p \frac{1}{p-1} \sum_{j=1, j \neq i}^p \text{cov}(X_i, X_j) = \sum_{i=1}^p \text{var}(T)$$

Luego, se tiene el resultado. Usando el Teorema 3.7 se tiene además que si los ítems son tau-equivalentes λ_2 también coincide con la confiabilidad. \square

Otra propiedad interesante del coeficiente α es que es la media de todos los posibles coeficientes de split-half [27]. Luego de cierta forma resuelve el problema de elegir las mitades para correlacionar del coeficiente de split-half.

5. λ_4 de Guttman: Este coeficiente se interpreta como el mayor coeficiente de split-half posible. Como en general es imposible computacionalmente encontrar exactamente las mitades de la prueba que maximizan al coeficiente split-half, existen diversas heurísticas para poder estimarlo. Por el teorema (3.5), este coeficiente debiera de ser una cota inferior para la confiabilidad, de hecho algunas personas la llaman la mayor cota inferior. Sin embargo Thompson, Green y Yag mostraron en [43] que además de poder subestimar la confiabilidad, también puede sobrestimarla por efectos de *capitalizing on chance*, es decir por utilizar propiedades aleatorias de la muestra a su favor.
6. λ_6 de Guttman: Vamos a denotar al coeficiente de correlación múltiple entre el ítem k con el resto de los ítems como $R_{k\bullet}^2$. Así el coeficiente λ_6 es la expresión de la confiabilidad de la ecuación (3.31), donde el término $\sum_{k=1}^p \text{var}(E_k)$ se reemplaza por $\sum_{k=1}^p (1 - R_{k\bullet}^2)$. Su expresión es:

$$\lambda_6 = \frac{\text{var}(X) - \sum_{k=1}^p (1 - R_{k\bullet}^2)}{\text{var}(X)} \quad (3.42)$$

Una forma de entender por qué este coeficiente coincide con la confiabilidad es mirar el modelo de TCT como un modelo de análisis factorial y considerar que los ítems son congenéricos. Esto se resume en los siguientes supuestos:

- (a) El primer supuesto es que existe sólo un factor común a todos los ítems que vamos a llamar T y se interpreta como el puntaje verdadero de TCT. Es decir se interpreta al puntaje verdadero como una variable que explica a la varianza común de los ítems.
- (b) En el modelo de análisis factorial se descompone la varianza de cada ítem en una varianza común a los otros ítems denotada por h_k^2 más una varianza específica L_k^2 . La varianza específica, a su vez se descompone en varianza única s_k^2 y varianza de error e_k^2 . Si se consideran variables estandarizadas se tiene la siguiente ecuación:

$$\text{var}(X_k) = 1 = h_k^2 + s_k^2 + e_k^2 \quad (3.43)$$

Entonces el segundo supuesto que se va a hacer, es que $s_k^2 = 0$, o sea que la varianza específica se debe sólo a varianza de error, esto implica que $1 - h_k^2 = e_k^2$

Si se cumplen los supuestos anteriores podemos establecer las siguientes relaciones entre el modelo de análisis factorial y el de teoría clásica:

- La varianza común o comunalidad h_k^2 de cada ítem representa la varianza del ítem producida por el puntaje verdadero
- La varianza de error de análisis factorial coincide con la de teoría clásica $e_k^2 = \text{var}(E_k)$.

Por otro lado en el capítulo de análisis factorial se vió que el coeficiente $R_{k\bullet}^2$ es una cota inferior para la comunalidad de un ítem:

$$R_{k\bullet}^2 < h_k^2 \quad (3.44)$$

Por lo tanto:

$$1 - R_{k\bullet}^2 > 1 - h_k^2 = e_k^2 = \text{var}(E_k) \quad (3.45)$$

Luego el coeficiente λ_6 consiste en remplazar $\sum_{k=1}^p \text{var}(E_k) = \sum_{k=1}^p e_k^2$ por $\sum_{k=1}^p (1 - R_{k\bullet}^2)$, y por la ecuación (3.45), se tiene que:

$$\lambda_6 \leq R_X \quad (3.46)$$

Sin embargo, si los ítems son multidimensionales el coeficiente λ_6 no es una cota inferior para la confiabilidad ya que en este caso:

$$\text{var}(E_k) > e_k^2 \quad (3.47)$$

La ecuación (3.47) se tiene porque en el caso de que se tengan más factores la varianza verdadera desde el punto de vista de teoría clásica, sería sólo la que toma en cuenta al primer factor. Sin embargo como existen más factores la varianza común o comunalidad podría ser mayor a la varianza verdadera. Es más en este caso λ_6 podría sobrestimar la confiabilidad.

7. ω_h : Este coeficiente usa el modelo de análisis factorial para determinar la varianza verdadera y la varianza de error. Hace los mismos supuestos del coeficiente λ_6 , en particular supone que los ítems son congénéricos. Si embargo en vez de estimar la comunalidad con el coeficiente de correlación múltiple, la estima aplicando un modelo de un factor. Esta diferencia es importante ya que en caso de multidimensionalidad no va a sobrestimar la confiabilidad como el coeficiente λ_6 sino que la va a subestimar.

Si consideramos el modelo de un factor, donde s.p.g. se asume que $\text{var}(T) = 1$, se tiene la siguiente ecuación:

$$X_i = \beta_i T + E_i \quad (3.48)$$

Donde

$$\text{var}(X_i) = \beta_i^2 + \text{var}(E_i) \quad (3.49)$$

$$\text{var}(X) = \left(\sum_{i=1}^p \beta_i \right)^2 + \sum_{i=1}^p \text{var}(E_i) \quad (3.50)$$

Luego se define ω_h como [28]:

$$\omega_h = \frac{(\sum_{i=1}^p \beta_i)^2}{\text{var}(X)} \quad (3.51)$$

Es importante notar, que al estimar la confiabilidad a partir de un modelo de análisis factorial, el supuesto teórico que se hace, que no estaba en teoría clásica, es que el puntaje verdadero es el factor que representa la varianza común entre todos los ítems. La estimación de este coeficiente se puede hacer por medio de modelos de análisis factorial confirmatorio.

Todos estos coeficientes, para coincidir con la confiabilidad necesitan al menos que los ítems sean congénéricos. Este supuesto es lo mismo que pedir que los ítems sean unidimensionales (donde el factor que los explicaría sería T). En el caso de pruebas con más dimensiones no son adecuados para medir la confiabilidad.

3.1.3. Errores

El error de medición es la variable $E = X - T$. Se define al error estándar de medición de un instrumento SEM como la desviación estándar del error de medición:

Definición 3.10 (Error estándar de medición SEM)

$$SEM = \text{sd}(E) \quad (3.52)$$

Utilizando los supuestos de teoría clásica de la ecuación (3.2), tenemos que:

$$SEM^2 = \text{var}(E) = \text{var}(X) - \text{var}(T) \quad (3.53)$$

Cómo $\text{var}(T) = \text{var}(X) \cdot R_X$, se tiene que:

$$SEM = \text{sd}(X) \sqrt{(1 - R_X)} \quad (3.54)$$

La ecuación (3.54) permite estimar al error estándar de medición usando un estimador de la confiabilidad y de la desviación estándar de X . Naturalmente, mientras menor sea la confiabilidad el error estándar de medición es mayor. La información que presenta el error estándar de medición es importante porque permite estimar un rango en el cual se mueven los puntajes verdaderos (considerando que los puntajes observados se distribuyen de forma normal) [15]. Poder establecer este rango es relevante, de hecho el primer estándar del capítulo de confiabilidad del libro *Standards for Educational and Psychological Testing* [1] dice “Para cada puntaje, subpuntaje o combinación de puntajes que vaya a ser interpretada, estimaciones de confiabilidad relevantes, errores estándares y errores estándares de medición o funciones de información de test deben ser reportadas”.

3.1.4. Estimación de la confiabilidad

Dentro de las formas de estimar confiabilidad, existen tres tipos principales:

1. Test-retest: Esta forma de estimación implica aplicar la prueba dos veces, donde las aplicaciones están separadas por un intervalo de tiempo que permita evitar el efecto de la memorización. Luego se considera cada aplicación como una forma paralela, por lo tanto la correlación de los puntajes entrega el estimador de confiabilidad. Esta correlación se considera como un coeficiente de confiabilidad, pero también de la estabilidad de la medición. Esta forma de estimar es muy criticada porque las personas tenderían a replicar sus respuestas anteriores [27].
2. Formas paralelas: En este caso se crean dos formas paralelas de una prueba y se calcula la confiabilidad como la correlación entre el puntaje de ambas.
3. Coeficientes de consistencia interna: En esta categoría están todos los coeficientes que estiman la confiabilidad de una prueba a partir de las relaciones entre sus ítems. Por esta razón sólo requieren una aplicación de la prueba a analizar. En el caso de split-half es necesario aplicar la prueba y elegir las dos mitades a correlacionar, en general se forma una mitad con los ítems pares y la otra con los impares.

Para estimar los coeficientes en el caso de λ_1 , λ_2 , α y λ_6 , hay que remplazar las varianzas, covarianzas o correlaciones de sus fórmulas por estimadores de estas cantidades. Para estimar el coeficiente λ_4 se utilizan diversas estrategias para encontrar las mitades de la prueba que más correlacionan, para más detalles ver [33]. Para estimar el coeficiente ω_h hay que hacer un modelo de análisis factorial usual o confirmatorio para estimar los coeficientes β_i y con ellos y la varianza del puntaje total calcular el coeficiente.

En general los dos primeros métodos de estimación de confiabilidad no son muy utilizados. En el caso de Test-retest porque además del esfuerzo de aplicar dos veces la prueba a las mismas personas, no genera estimaciones confiables. Por otro lado el hacer dos pruebas paralelas no es muy común, ya que para que ambas formas sean paralelas es necesario que cumplan muchas propiedades y es mejor usar ese esfuerzo en crear una prueba más larga y confiable.

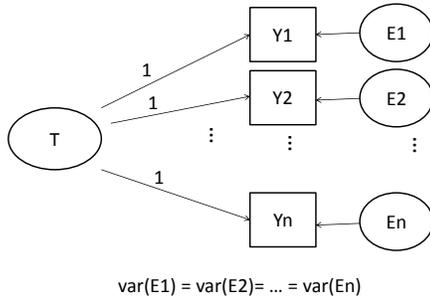
Luego los coeficientes más utilizados son los que necesitan sólo una aplicación de la prueba. El coeficiente más usado es el α de Cronbach a pesar de no ser el mejor estimador. Por ejemplo λ_2 es siempre una cota inferior para la confiabilidad mayor que α y en simulaciones también presenta mejores propiedades [41]. Otro punto importante respecto a la elección del coeficiente son las hipótesis que deben de cumplir los ítems para que el coeficiente coincida con la confiabilidad. En particular el coeficiente α pide el supuesto de tau-equivalencia que es muy estricto. A diferencia del coeficiente ω_h que pide sólo el supuesto de ítems congenéricos para coincidir con la confiabilidad, supuesto que puede probarse a través de un análisis factorial a los ítems. Existen muchos papers que critican el uso del coeficiente α y proponen calcular la confiabilidad con ideas similares a las del coeficiente ω_h , como por ejemplo [24], [38], [47], [14], [6].

3.1.5. Testeo de las hipótesis de teoría clásica con análisis factorial confirmatorio

Las hipótesis de ítems paralelos, tau-equivalentes o congenéricos se pueden testear a través del análisis factorial confirmatorio. Los modelos para testear se presentan en la Figura 3.1 [14]. Todos estos modelos asumen unidimensionalidad, luego es importante verificar esta propiedad en los ítems con un análisis factorial primero. Si la unidimensionalidad no se tiene, todos los coeficientes de consistencia interna van a dar resultados imprecisos [14].

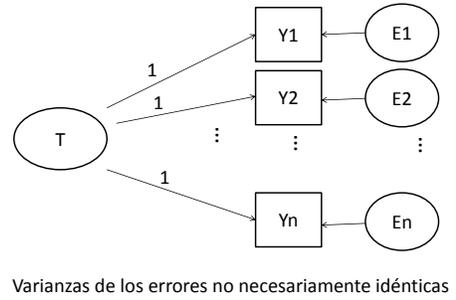
Los modelos van en orden creciente de complejidad, luego siempre el modelo con ítems congenéricos va a ser el que mejor estime a los datos. Por esta razón para seleccionar el modelo es importante ver su ajuste general pero también su parsimonia, si no siempre se va a elegir al menos restrictivo (ítems congenéricos). Hay dos formas de tomar en cuenta esto, una es determinar si la disminución del estadístico chi-cuadrado entre un modelo menos restrictivo que otro es significativa. Aunque el estadístico de chi-cuadrado es muy dependiente del tamaño muestral, en este caso no es importante ya que ambos modelos tienen la misma cantidad de datos. La otra forma es analizar índices de ajuste que toman en cuenta la parsimonia del modelo como el criterio de información de Akaike.

Modelo 1: Ítems Paralelos



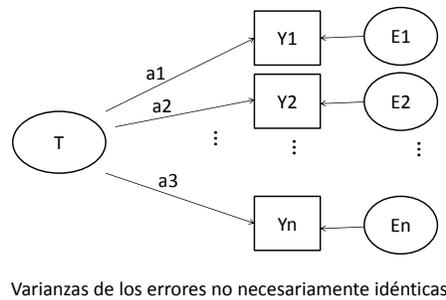
(a)

Modelo 2: Ítems tau-equivalentes



(b)

Modelo 3: Ítems congénéricos



(c)

Figura 3.1: Modelos de análisis factorial confirmatorio para testear las hipótesis: Los ítems son paralelos; los ítems son tau-equivalentes o los ítems son congénéricos

3.1.6. Estadísticos de los ítems

Para describir a cada ítem que forma una prueba, existen dos estadísticos importantes:

1. Dificultad: Es el porcentaje de personas que respondieron correctamente a cada ítem. Aunque este estadístico cumple que a valores mas altos el ítem es más fácil tradicionalmente se le ha llamado dificultad [28].
2. Discriminación: Es una medida de cuanto discrimina el ítem entre las personas con alto y bajo puntaje. Se calcula como la correlación entre el ítem X_i y el puntaje total X . Usualmente se toma un puntaje corregido donde no se considera el ítem, es decir $X^* = X - X_i$ para que la correlación del ítem consigo mismo no genere una sobrestimación de la discriminación.

Hay dos formas de calcular la correlación entre el ítem y el puntaje corregido, una es la correlación usual de Pearson y la otra es la correlación biserial, esta se define como:

Definición 3.11 (Correlación Biserial) *Sea X_i una variable que representa un ítem dicotómico. Se supone que existe una variable continua \tilde{X}_i que define a la variable X_i*

por medio de la siguiente relación.

$$X_i = \begin{cases} 1 & \text{si } \tilde{X}_i > t_i, \\ 0 & \text{si } \tilde{X}_i \leq t_i. \end{cases} \quad (3.55)$$

Si X es el puntaje de la prueba (total o corregido por el ítem i), entonces se define:

$$\text{Correlación biserial entre } X_i \text{ y } X = \text{cor}(\tilde{X}_i, X) \quad (3.56)$$

En general lo que se espera de los ítems es que sean lo más discriminantes posibles, y ojalá tengan un rango amplio de dificultades. Aunque la distribución de las dificultades que se necesite puede depender del objetivo del test, por ejemplo si el objetivo es distinguir entre personas que se espera que tengan un alto desempeño, va a ser mejor utilizar ítems difíciles, en el caso en que se quiera evaluar a personas con bajo desempeño se debieran de seleccionar ítems más fáciles. Por otra parte si se quiere una prueba de diagnóstico, conviene tener ítems que abarquen un rango amplio de dificultades.

3.2. Teoría clásica de test aplicada a la prueba CAM

A continuación se presenta la aplicación de teoría clásica de test en las pruebas CAM total y depurada. Primero se mostrará una descripción estadística de los puntajes observados y luego un análisis de confiabilidad y de la dificultad y discriminación de los ítems.

3.2.1. Estadísticos descriptivos

Se presentan los estadísticos descriptivos de cada prueba en la Tabla 3.1.

Estadísticos	61 ítems	28 ítems
Media	22,9	12,32
Mediana	24	13
sd	7,60	5,02
Mín	0-4	1
Máx	42	25

Tabla 3.1: Estadísticos descriptivos del puntaje de la prueba CAM total (61 ítems) y de la prueba CAM depurada (28 ítems)

Se presentan también los histogramas del puntaje total en la prueba CAM y en la prueba CAM depurada en la Figura 3.2.

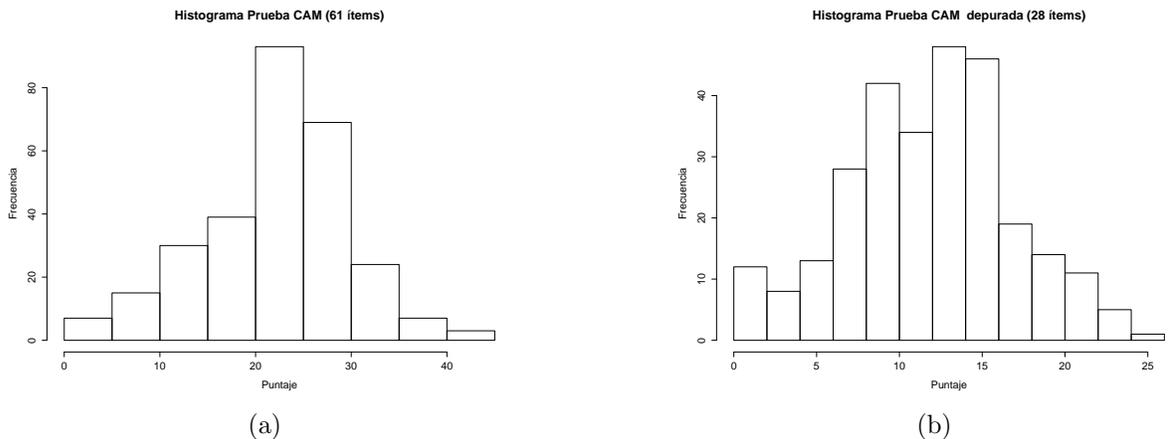


Figura 3.2: Histograma de los puntajes de la prueba CAM total y de la prueba CAM depurada.

3.2.2. Confiabilidad

Se testeó si se cumplían las hipótesis de ítems paralelos, tau-equivalentes o congénicos aplicando los modelos de la Figura 3.1 en cada una de las dos pruebas. Los índices de ajuste para los modelos aplicados a la prueba CAM total se presentan en la Tabla 3.2 y para los modelos aplicados sobre la prueba depurada en la Tabla 3.3.

Estadísticos	Ítems paralelos	Ítems equivalentes	tau-	Ítems congénéricos
χ^2	-	3955,4		3715
<i>g.l.</i>	-	1829		1769
RMSEA	-	0,064*		0,062*
SRMR	-	0,089		0,078
GFI	-	0,68		0,69
AIC	-	4079,4		3898
Comparación	-	Ítems paralelos vs tau-equivalentes		Ítems equivalentes vs congénéricos
$\chi_{i+1}^2 - \chi_i^2$	-	-		240,43
<i>g.l.</i> _{<i>i+1</i>} - <i>g.l.</i> _{<i>i</i>}	-	-		60
<i>p - valor</i>	-	-		<0.001

Tabla 3.2: Estadísticos de ajuste en los modelos para testear las hipótesis de ítems paralelos, tau-genéricos y congénéricos en la prueba CAM total (61 ítems)

Se marcaron con un asterisco * los índices que representaban un ajuste aceptable (ninguno mostró un ajuste excelente) [35]. En la prueba CAM total no se pudo ajustar el modelo de ítems paralelos, ya que era necesario imponer varianzas negativas.

Respecto a la calidad de los modelos en la prueba CAM total, por el criterio de chi-cuadrado y por el criterio de Akaike, se tiene que es mejor el de ítems congénéricos. También todos los índices de ajuste muestran que el modelo congénérico es mejor. Sin embargo este modelo sólo describe los datos de manera aceptable, una explicación para esto es que, como se vio en el análisis factorial exploratorio, la prueba CAM total era multidimensional. Es más otra razón de porque el modelo de ítems congénéricos sobre la prueba total no fue muy bueno es porque el coeficiente que relaciona al ítem 5 con el puntaje verdadero fue negativo, lo cual muestra que el modelo no se ajustó a lo esperado teóricamente en ese ítem.

Respecto a los resultados en la prueba CAM depurada el ajuste fue aceptable en los tres modelos. Sin embargo el estadístico de chi-cuadrado indicó que no hay una mejora importante entre considerar ítems paralelos o ítems tau-equivalentes, pero sí entre considerar ítems tau-equivalentes o congénéricos. Según el criterio de Akaike el mejor modelo es el de ítems paralelos y según la comparación del chi-cuadrado, el mejor modelo es el de ítems congénéricos.

Si se compara el ajuste de los modelos entre ambas pruebas, el ajuste de los tres modelos fue mejor en la prueba depurada que en la total. Esto se explica por que en el análisis factorial hecho en el capítulo anterior, se vio que la prueba depurada es realmente unidimensional a diferencia de la total.

Estos resultados nos permiten interpretar mejor los estimadores de la confiabilidad de la Tabla 3.4.

En general, podemos concluir que la confiabilidad en el caso de la prueba CAM total está

Estadísticos	Ítems paralelos	Ítems equivalentes	tau-	Ítems congénéricos
χ^2	707,26	698,46		644,86
<i>g.l.</i>	404	377		350
RMSEA	0,052*	0,055*		0,054*
SRMR	0,077*	0,077*		0,066*
GFI	0,847	0,848		0,859
AIC	711,25	756,48		756,86
Comparación	-	Ítems paralelos vs tau-equivalentes		Ítems equivalentes vs congénéricos
$\chi_{i+1}^2 - \chi_i^2$	-	8,781		53,619
<i>g.l.</i> _{<i>i+1</i>} - <i>g.l.</i> _{<i>i</i>}	-	27		27
<i>p - valor</i>	-	0,99		<0,001**

Tabla 3.3: Estadísticos de ajuste en los modelos para testear las hipótesis de ítems paralelos, tau-genéricos y congénéricos en la prueba CAM depurada (28 ítems)

Coeficiente	Confiabilidad		SEM	
	CAM total	CAM depurada	CAM total	CAM depurada
λ_1	0,77	0,75	-	-
λ_2	0,8	0,78	3,4	2,35
α o λ_3	0,79	0,78	3,48	2,35
λ_4	0,8	0,87	3,4	1,81
λ_6	0,86	0,81	2,84	2,19
ω_h	0,79	0,78	3,48	2,35

Tabla 3.4: Coeficientes de confiabilidad y errores estándares de medición en ambas pruebas

entre 0,79 y 0,8, ya que por ser multidimensional (en el capítulo de análisis factorial se vio que se necesitaban al menos 5 factores para describirla) el coeficiente λ_6 estaría sobrestimando la confiabilidad. Sin embargo por la naturaleza multidimensional de esta prueba, ninguno de estos coeficientes puede medir de forma precisa la confiabilidad.

Por otra parte, en la prueba CAM depurada la confiabilidad varía entre 0,78 y 0,81. Esto porque los resultados de los modelos de análisis factorial confirmatorio muestran que los coeficientes λ_2 , α , λ_6 y ω_h no debieran de diferir demasiado del valor real de la confiabilidad. Sin embargo sabemos que el coeficiente λ_4 puede sobrestimar la confiabilidad por efectos de *capitalizing on chance*, por esto no lo tomamos en cuenta para concluir.

La estimación del error de medición depende del coeficiente de confiabilidad que se utilice, por esto para cada estimación de confiabilidad se estimó el error estándar de medición.

En conclusión la prueba depurada, a pesar de tener muchos menos ítems que la prueba total, tiene casi la misma confiabilidad. Esto muestra que la selección realmente fue un aporte ya que se está midiendo el mismo constructo, en menos tiempo y con un nivel de confiabilidad bueno [37].

3.2.3. Estadísticos de los ítems: Dificultad y Discriminación

La estimación de la dificultad de cada ítem en la prueba CAM total y la prueba CAM depurada se presentan en las Tablas 3.5 y 3.6.

ítem	dificultad	ítem	dificultad	ítem	dificultad	ítem	dificultad
I18d	0,73	I24c	0,44	I28b	0,36	I17d	0,22
I16a	0,72	I13d	0,44	I19c	0,36	I20b	0,22
I19b	0,69	I17a	0,44	I20e	0,36	I13c	0,22
I24a	0,68	I1c	0,44	I23	0,35	I10d	0,2
I2	0,66	I10b	0,43	I10c	0,33	I18b	0,2
I14	0,62	I1a	0,42	I26b	0,31	I28a	0,18
I4	0,61	I13a	0,42	I18c	0,29	I20d	0,17
I16b	0,59	I18a	0,41	I11	0,29	I8	0,15
I24b	0,59	I3	0,41	I10a	0,28	I9	0,13
I19a	0,53	I26d	0,41	I7	0,26	I5	0,12
I6	0,53	I28c	0,41	I15	0,26	I22	0,12
I17c	0,51	I26c	0,4	I20a	0,24	I1b	0,11
I26a	0,48	I16c	0,4	I27	0,23	I13b	0,03
I21	0,48	I16d	0,4	I28d	0,23		
I10e	0,47	I17b	0,4	I25	0,23		
I28e	0,47	I24d	0,39	I20c	0,22		

Tabla 3.5: Dificultades de los ítems de la prueba CAM total

ítem	dificultad	ítem	dificultad	ítem	dificultad	ítem	dificultad
I18d	0,74	I21	0,49	I26d	0,42	I23	0,36
I16a	0,73	I10e	0,48	I28c	0,42	I18c	0,3
I19b	0,7	I28e	0,48	I26c	0,41	I15	0,27
I24a	0,69	I24c	0,45	I17b	0,41	I27	0,24
I24b	0,6	I17a	0,45	I24d	0,4	I25	0,23
I19a	0,54	I10b	0,43	I28b	0,37	I17d	0,22
I26a	0,49	I18a	0,42	I19c	0,36	I28a	0,19

Tabla 3.6: Dificultades de los ítems de la prueba CAM depurada

Se presentan en la Figura 3.3 los histogramas de las dificultades de los ítems de la prueba CAM y de la selección de ítems de la prueba CAM depurada. Esto para ver como se distribuye la dificultad de los ítems en cada prueba.

De la Figura 3.3 se puede concluir que las dificultades de la prueba CAM total se concentran en el rango de 0,4 a 0,5 y que hay una mayor presencia de ítems difíciles que fáciles. En la prueba CAM depurada también se presenta una concentración de dificultades en el rango de 0,4 a 0,5, sin embargo la proporción entre ítems difíciles y fáciles es más pareja. Por otra parte se tiene que los ítems con dificultades extremadamente bajas (es decir los demasiado fáciles) no formaron parte de la prueba CAM depurada. Además el ítem más

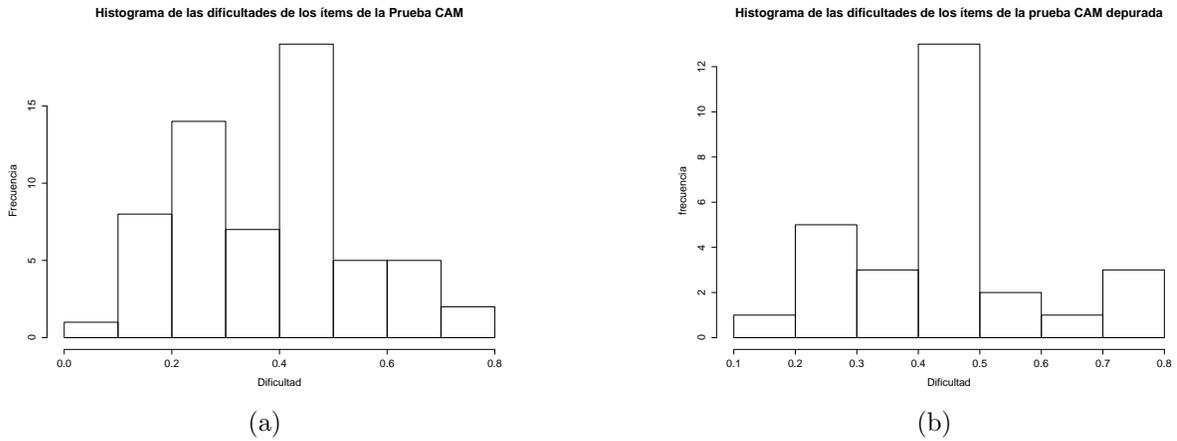


Figura 3.3: Histograma de las dificultades de los ítems de la prueba CAM total y de la prueba CAM depurada

difícil en la prueba CAM depurada fue respondido correctamente por un 19% de las personas, a diferencia del más difícil de la prueba CAM, que fue respondido por sólo un 3% de las personas.

Se presentan las discriminaciones de los ítems de las pruebas CAM y CAM depurada en las Tablas 3.7 y 3.8. Estas discriminaciones se calcularon como la correlación de Pearson entre el ítem y el puntaje corregido.

Es importante notar que los ítems que tuvieron discriminaciones bajas en teoría clásica coinciden bastante con los ítems que fueron eliminados a partir del análisis factorial. Esto es entendible porque la mayoría de los coeficientes que definen al factor que representaba al *conocimiento de alumnos y matemática* son significativamente positivos y un resto menor es cercano a 0 (ver Sección 2.5 del capítulo de análisis factorial). Luego este factor es similar a la suma de los ítems. Por esto es natural que los ítems que en teoría clásica mostraron un comportamiento no muy bueno, en análisis factorial también. Es bueno notar que hay una coherencia entre los resultados de análisis factorial y de teoría clásica de test, sin embargo el análisis factorial nos permitió interpretar de manera mejor los índices de confiabilidad.

Se puede notar también que en la selección, a diferencia de la prueba total, todos los ítems se correlacionan de manera alta con la escala.

ítem	discriminación	ítem	discriminación	ítem	discriminación	ítem	discriminación
I28e	0,42	I19c	0,31	I1c	0,2	I28d	0,1
I16a	0,4	I24b	0,3	I17a	0,2	I10c	0,1
I19b	0,4	I20b	0,29	I15	0,2	I1a	0,1
I19a	0,38	I13d	0,29	I17d	0,19	I16b	0,1
I24a	0,37	I18a	0,29	I2	0,19	I4	0,09
I27	0,36	I23	0,26	I8	0,19	I1b	0,09
I10b	0,34	I20c	0,23	I3	0,18	I9	0,08
I25	0,34	I20a	0,23	I7	0,18	I26b	0,07
I10e	0,33	I17b	0,23	I13c	0,18	I11	0,05
I21	0,33	I18c	0,23	I20d	0,17	I13b	0,05
I26a	0,32	I24c	0,22	I16c	0,15	I10a	0,03
I13a	0,32	I20e	0,22	I14	0,15	I18b	0,01
I18d	0,32	I17c	0,22	I22	0,13	I5	-0,01
I28b	0,31	I28a	0,21	I10d	0,12		
I26c	0,31	I26d	0,21	I16d	0,12		
I28c	0,31	I24d	0,2	I6	0,11		

Tabla 3.7: Discriminaciones de los ítems de la prueba CAM total

ítem	discriminación	ítem	discriminación	ítem	discriminación	ítem	discriminación
I28e	0,45	I28c	0,36	I10b	0,29	I24c	0,24
I27	0,41	I19c	0,35	I10e	0,26	I24d	0,22
I19b	0,38	I19a	0,34	I18c	0,25	I17d	0,22
I24a	0,37	I28b	0,32	I23	0,25	I28a	0,22
I25	0,37	I16a	0,31	I18a	0,25	I17b	0,19
I26a	0,37	I21	0,3	I18d	0,25	I17a	0,17
I26c	0,37	I24b	0,3	I26d	0,25	I15	0,17

Tabla 3.8: Discriminaciones de los ítems de la prueba CAM depurada

Capítulo 4

Teoría de Respuesta al Ítem

La Teoría de Respuesta al Ítem (TRI) esta formada por un conjunto de modelos que permiten representar y caracterizar el comportamiento de ítems de una prueba y de personas que la responden. También permite entender la relación entre estos comportamientos. Su principal fortaleza es que las estimaciones de los parámetros de los ítems y las personas no dependen de la muestra (si es que se cumplen las hipótesis del modelo). Esta propiedad es muy importante y es una de las principales ventajas de la TRI sobre la teoría clásica de test. En teoría clásica los parámetros de dificultad de los ítems dependen de la muestra que se use. Esto es problemático porque para poder determinar el nivel de habilidad que tienen los individuos evaluados es necesario conocer cuan difícil es la prueba. Sin embargo la dificultad de la prueba depende de las dificultades de sus ítems, las que a su vez dependen de como los participantes respondieran. Por lo tanto la teoría clásica se enfrentaba al problema de que la medición depende del objeto medido [29].

En términos matemáticos la teoría de respuesta al ítem es un conjunto de modelos no lineales que se basan en describir la probabilidad de responder correctamente a un ítem como una función de una variable que representa el nivel de habilidad del individuo y parámetros propios del ítem. Si consideramos que la habilidad o constructo que mide la prueba es una variable continua θ que varía entre los individuos, tenemos que el modelo de la TRI propone la siguiente ecuación para la probabilidad de que una persona con habilidad θ responda correctamente al ítem i .

$$\mathbb{P}_i(\theta) = f(\theta|b_i, a_i, c_i) \quad (4.1)$$

En donde b_i , a_i y c_i son los parámetros de dificultad, discriminación y de respuesta al azar respectivamente y todos dependen del ítem. En todos los modelos de IRT está presente el parámetro de dificultad, en cambio a_i y c_i no siempre. Estos parámetros permiten distinguir los ítems que aportan de los que no en una prueba y seleccionarlos según el objetivo específico que se tenga. Respecto a la función f , lo usual es definirla cómo una variación de la función logística (ver ecuación (4.2)) o de la ojiva normal (ver ecuación (4.3)).

Definición 4.1 (Función Logística)

$$\Psi(x) = \frac{e^x}{1 + e^x} \quad (4.2)$$

Definición 4.2 (Función Ojiva Normal)

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (4.3)$$

Ambas funciones son muy parecidas, de hecho Haley [17] demostró que

$$|\Phi(x/1,7) - \Psi(x)| < 0,01 \quad \forall x \quad (4.4)$$

Como la función logística es más fácil de calcular computacionalmente, y es casi igual a la ojiva normal, la mayoría de los softwares y libros del tema utilizan esta función. En particular en esta memoria para presentar los métodos de estimación y para los cálculos que se realicen sólo se utilizarán los modelos de IRT con la función logística, sin embargo en algunos teoremas y resultados teóricos se va a usar el modelo de la ojiva normal.

Respecto a si existe una explicación de porqué se elige la función logística o normal, Lord [26] dice que “No convincing a priori justification exists [...]. The model must be justified on the basis of the results obtained, not a priori grounds”¹.

Con las definiciones ya vistas, se pueden describir los tres modelos principales que usa la TRI para representar la probabilidad de una respuesta correcta. El uso de uno u otro modelo depende de las características particulares de la prueba que se quiera modelar y de las necesidades del investigador. Estos modelos son:

1. Modelo de Rasch

$$\mathbb{P}_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} \quad (4.5)$$

En este modelo la probabilidad de responder correctamente un ítem sólo depende de la dificultad del ítem y del nivel de habilidad del individuo. A pesar de ser el modelo más simple es muy utilizado, ya que la dificultad del ítem y la habilidad del individuo, están en la misma escala. Luego se pueden relacionar a los ítems y a las personas por medio de los gráficos ítem-persona, lo que entrega una riqueza interpretativa importantísima.

2. Modelo de dos parámetros 2PL

$$\mathbb{P}_i(\theta) = \frac{e^{(a_i(\theta-b_i))}}{1 + e^{(a_i(\theta-b_i))}} \quad (4.6)$$

En este modelo se agrega el parámetro de discriminación a_i , este permite diferenciar los ítems que separan más o menos a las personas según su nivel de habilidad. Este modelo es muy interesante ya que tiene una relación con el análisis factorial y permite identificar los ítems que mejor distinguen entre las personas con más o menos nivel de habilidad.

3. Modelo de 3 parámetros 3PL

$$\mathbb{P}_i(\theta) = c_i + (1 - c_i) \frac{e^{(a_i(\theta-b_i))}}{1 + e^{(a_i(\theta-b_i))}} \quad (4.7)$$

¹En el mismo libro donde está esta cita, el autor deduce el modelo usando la ojiva normal, sin embargo hace tantos supuestos que no lo considera como una deducción adecuada para justificar al modelo en general

Este modelo además de incorporar al parámetro de discriminación, agrega el parámetro de responder al azar c_i . Este buscaría modelar las respuestas correctas que se dan porque las personas respondieron al azar y no por su nivel de habilidad.

En este capítulo sólo se van a describir los modelos de Rasch y de dos parámetros. El modelo de tres parámetros en el caso de la prueba CAM no se justifica, ya que las personas evaluadas no tenían motivos a priori para responder al azar. Además el modelo de tres parámetros tiene mayores problemas de estimación que el de Rasch y el de dos parámetros.

En lo que sigue se presentan primero los supuestos y propiedades que son comunes a todos los modelos. Después se describe brevemente el modelo de Rasch y el de IRT a dos parámetros. Por último se presenta la aplicación de estos dos modelos a la prueba CAM total y depurada.

La notación de este capítulo es similar a la de los anteriores, se va a considerar que se tiene una prueba de p ítems, y una muestra de N personas. Definimos x_i a la variable que vale 1 si el ítem i se respondió correctamente y 0 si no. Al vector que contiene a las variables de todos los ítems es $x = (x_1, \dots, x_p)$ vector multivariado que varía entre las personas.

4.1. Supuestos de los modelos

En general en las teorías que buscan describir al comportamiento humano a partir de variables latentes (habilidades) se supone que hay un espacio de m variables, que determinan las respuestas de un individuo a un ítem [18]. En el caso de la TRI el primer supuesto que se hace es que ese espacio tiene dimensión $m = 1$ y que la variable que lo forma es la habilidad θ . Este supuesto de unidimensionalidad es muy fuerte ya que implica que la única característica del individuo que explica sus respuestas a los ítems es su nivel en la habilidad θ .

El segundo supuesto del modelo es el de independencia local de los ítems. Aunque la unidimensionalidad implica independencia local es bueno conceptualizarlo aparte para considerarlo al diseñar pruebas. Además se puede tener independencia local con datos no unidimensionales [29]. La independencia local de los ítems se refiere a que para un nivel de habilidad dado, la probabilidad de responder correctamente un ítem no dependa de las respuestas a los otros ítems. Matemáticamente el supuesto de independencia local se escribe como [26]:

$$\mathbb{P}(x_i = 1|\theta) = \mathbb{P}(x_i = 1|\theta, x_j, x_k, \dots) \quad i \neq j, k, \dots \quad (4.8)$$

Este supuesto es muy importante, ya que permite escribir analíticamente la verosimilitud y gracias a eso se puede hacer estimación por máxima verosimilitud. Es importante notar que este supuesto no significa que los ítems tengan correlación 0 entre sí, sino que la correlación entre ellos es producida porque miden la misma variable latente θ , y si se saca el efecto de esta variable la correlación debiera de ser 0.

El último supuesto del modelo es que $\mathbb{P}_i(\theta)$ se puede describir con la función logística u ojiva normal.

Una condición suficiente, pero no necesaria para tener unidimensionalidad, independencia local y curvas características de los ítems con forma de ojiva normal es que la matriz de correlaciones tetracóricas tenga sólo un factor común para todos los ítems. Este resultado es de Lord [27] y es parte de un teorema muy importante que permite determinar cuanto se cumple el supuesto de unidimensionalidad. El teorema es el siguiente:

Teorema 4.3 *Si se cumple que:*

1. *Las frecuencias que constituyen la distribución multivariada de (x_1, \dots, x_p) para un grupo de examinados vienen de la dicotomización de p variables continuas (x_1^*, \dots, x_p^*) que se distribuyen de forma normal multivariada.*
2. *Las p variables continuas (x_1^*, \dots, x_p^*) tienen sólo un factor común.*

Entonces se tiene que para ese grupo de examinados la distribución multivariada de las variables dicotómicas (x_1, \dots, x_p) cumple las siguientes propiedades

- i) El espacio latente es unidimensional.*
- ii) Las curvas características del ítem son ojivas normales.*
- iii) Se tiene independencia local para los p ítems.*
- iv) La variable latente θ se distribuye de forma normal.*

El Teorema 4.3 es importante porque da una condición suficiente para determinar que las hipótesis del modelo se cumplen. Aunque se han inventado diversos índices y métodos para determinar si estas las hipótesis se cumplen el análisis factorial, que se justifica gracias a este teorema, es el más usado [29]. Como con datos reales es imposible tener una unidimensionalidad perfecta, es decir que un factor explique exactamente los datos, la unidimensionalidad es un asunto de grado, en donde mientras más varianza explique el primer factor, más unidimensional va a ser el conjunto de datos [29]. Es más, Lord [26] propone hacer un análisis factorial de la matriz de correlaciones tetracóricas y verificar que el primer valor propio sea “notablemente” superior al segundo y que el segundo valor propio no difiera mucho del resto, con esto los ítems se pueden considerar aproximadamente unidimensionales.

4.2. El modelo de Rasch

El modelo de Rasch propone que la probabilidad de que una persona con habilidad θ responda correctamente un ítem de dificultad b_i está dada por la siguiente ecuación:

$$\mathbb{P}_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} \quad (4.9)$$

La curva formada al graficar $\mathbb{P}_i(\theta)$ vs θ se llama curva característica del ítem. La dificultad se puede entender como un parámetro de locación de la curva. De hecho, gracias a la forma de la función logística la dificultad coincide con el valor de θ para el cual $\mathbb{P}_i(\theta) = 1/2$.

Si un ítem es más difícil que otro su curva va a estar más a la derecha, en ese caso para tener una probabilidad alta de responder hay que tener una mayor habilidad. Esto se puede ver en las Figuras 4.1a y 4.1b, que muestran la curva característica del ítem más fácil de la prueba CAM depurada (18d) y del más difícil (28a) respectivamente.

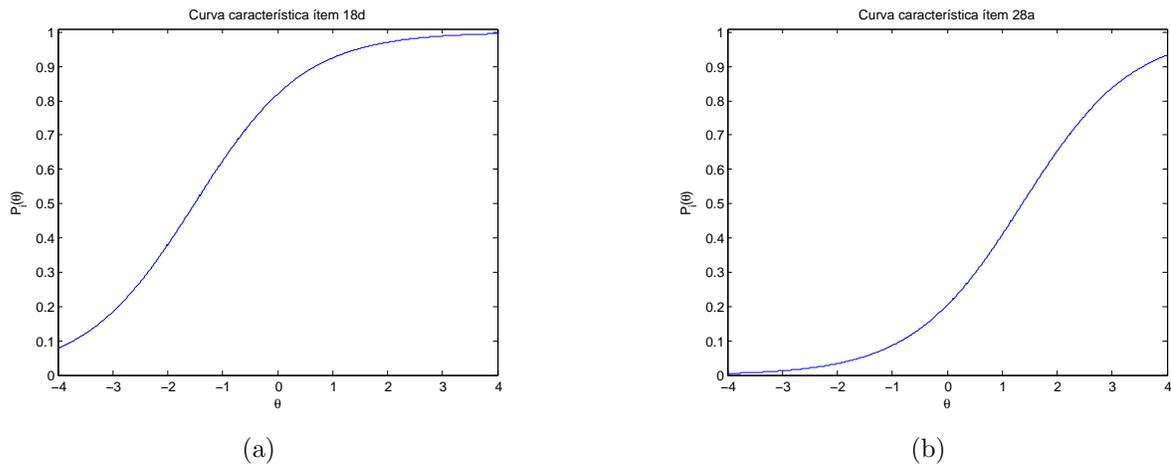


Figura 4.1: Curvas características del ítem para el ítem más fácil (a) y difícil (b) de la prueba CAM depurada.

Una propiedad muy importante del modelo de Rasch, es que si se usa la función logística para describir $\mathbb{P}_i(\theta)$, el puntaje total $\sum_{i=1}^p x_i$ es un estadístico suficiente para la habilidad θ [26]. Este resultado facilita mucho la estimación de habilidad especialmente en grupos muy grandes ya que se pueden agrupar las personas con el mismo puntaje [18].

4.2.1. Mapas ítems-personas

Una propiedad muy importante del modelo de Rasch es que gracias a la ecuación (4.9) la dificultad de los ítems y las personas están en la misma escala. Esto no se tiene en el modelo de dos parámetros ni en el de tres. Esto permite hacer los mapas ítems-personas, los cuales están formados por dos histogramas paralelos, uno de los ítems y otro de las personas. Estos histogramas permiten:

1. Determinar si la prueba fue fácil o difícil para la muestra, para hacer esto basta ver cual distribución está más arriba en la escala habilidad-dificultad.
2. Se puede medir la información que da cada ítem para cada habilidad (que se denota como $I_i(\theta)$). Para cada ítem la habilidad en la cual $I_i(\theta)$ toma su valor máximo es $\theta^* = b_i$. Luego como en el gráfico el nivel de habilidad coincide con el de dificultad se puede determinar para que personas hay más o menos ítems que dan el máximo de información.
3. Se puede determinar en que rangos de habilidad hacen falta ítems o que ítems están dando información de menos personas

Un ejemplo de mapa ítems-personas se muestra en la parte de resultados de este capítulo.

4.2.2. Ajuste de las dificultades y habilidades en el modelo de Rasch

En los modelos de la TRI no hay un consenso respecto a como medir el ajuste del modelo. En particular esto se debe a que la mayoría de los índices se basan en pruebas de chi-cuadrado, las cuales son muy dependientes de la cantidad de personas de la muestra [18].

Sin embargo en el modelo de Rasch, a diferencia de los otros modelos, se han popularizado índices de ajuste que permiten determinar la calidad del modelo en cada ítem y cada persona. Estos índices se llaman Infit y Outfit, para definirlos vamos a considerar que los datos de las N personas que respondieron los p ítems están codificados en una matriz de dimensión $N \times p$ denotada X , y vamos a utilizar la siguiente notación:

1. Valor observado: X_{ji}
2. Esperanza de X_{ji} : $\mathbb{E}(X_{ji}) = \mathbb{P}_i(\theta_j)$
3. Varianza de X_{ji} : $V_{ji} = \text{var}(X_{ji}) = \mathbb{P}_i(\theta_j) \cdot (1 - \mathbb{P}_i(\theta_j))$
4. Residuo: $y_{ji} = X_{ij} - \mathbb{P}_i(\theta_j)$
5. Residuo estandarizado: $z_{ji} = \frac{y_{ji}}{\sqrt{\mathbb{P}_i(\theta_j) \cdot (1 - \mathbb{P}_i(\theta_j))}}$

Con estas definiciones previas, se tiene que [7]:

Definición 4.4 (Outfit del ítem i)

$$Outfit = \sum_{j=1}^N \frac{z_{ji}^2}{N} \quad (4.10)$$

Definición 4.5 (Outfit de la persona j)

$$Outfit = \sum_{i=1}^p \frac{z_{ji}^2}{p} \quad (4.11)$$

Definición 4.6 (Infit del ítem i)

$$Infit = \sum_{j=1}^N \frac{z_{ji}^2 \cdot V_{ji}}{\sum_{j=1}^N V_{ji}} \quad (4.12)$$

Definición 4.7 (Infit de la persona j)

$$Infit = \sum_{i=1}^p \frac{z_{ji}^2 \cdot V_{ji}}{\sum_{i=1}^p V_{ji}} \quad (4.13)$$

Para poder interpretar estos estadísticos en primer lugar vamos a llamar valores extremos a los ítems en donde el porcentaje de personas que respondió correctamente fue muy cercano a 0, o muy cercano a 1. También vamos a llamar valores extremos a las personas cuyo porcentaje de respuestas correctas fue muy cercano a 0 o muy cercano a 1.

En el caso de los Outfit, estos son fuertemente influenciados por respuestas inesperadas que ocurren en ítems o personas que tienen valores extremos.

En cambio los Infit están ponderados por la información de cada ítem (que coincide con la varianza) [46]. Como la varianza de una variable dicotómica es $p(1-p)$, y tiene su máximo en $p = 1/2$, el Infit pondera más a los ítems o a las personas que tienen valores menos extremos, o donde la proporción de respuestas o personas que respondieron correctamente es más cercana a $1/2$, luego en este caso los valores extremos de ítems o personas son menos influyentes.

Simulaciones han mostrado que la suma de los z_{ji}^2 sobre los ítems o las personas se distribuye aproximadamente como chi-cuadrado con $p-1$, grados de libertad para una persona y $N-1$ grados de libertad para un ítem [46]. Además la esperanza del Infit y del Outfit es igual a 1.

A priori valores pequeños de estos índices indican que se tiene una mejor predicción. Sin embargo la esperanza del estadístico no es 0, si no que 1. Luego bajas muy fuertes del valor 1 dan cuenta de un sobre ajuste. De hecho ajustes cercanos a 0 pueden indicar una falla en el supuesto de independencia local. Por otro lado valores por sobre 1 dan cuenta de un mal ajuste [7]. Los rangos aceptables para estos estadísticos que se usan en la literatura son [7]:

Tipo de Test	Rango
Prueba con preguntas de elección múltiple de alto impacto	0,8-1,2
Prueba con preguntas de elección múltiple de impacto medio	0,7-1,3

Tabla 4.1: Rangos de ajuste para los estadísticos Infit y Outfit, tomados de [7].

4.3. El modelo de dos parámetros

El modelo de dos parámetros propone que la probabilidad de que una persona con habilidad θ responda correctamente un ítem de dificultad b_i y discriminación a_i está dada por la siguiente ecuación:

$$\mathbb{P}_i(\theta) = \frac{e^{(a_i(\theta-b_i))}}{1 + e^{(a_i(\theta-b_i))}} \quad (4.14)$$

La curva formada al graficar $\mathbb{P}_i(\theta)$ vs θ , al igual que en el modelo de Rasch, se llama curva característica del ítem. En este modelo también la dificultad es un parámetro de locación de la curva y coincide con el valor de θ para el cual $\mathbb{P}_i(\theta) = 1/2$.

El parámetro de discriminación determina si la pendiente de la curva en $\theta = b_i$ es más o menos empinada. Esto porque para este modelo $\frac{\partial \mathbb{P}}{\partial \theta}(b_i) = \frac{a_i}{4}$.

Se presentan en la Figura 4.2 los gráficos de dos curvas características de ítems con dificultades similares y distintas discriminaciones.

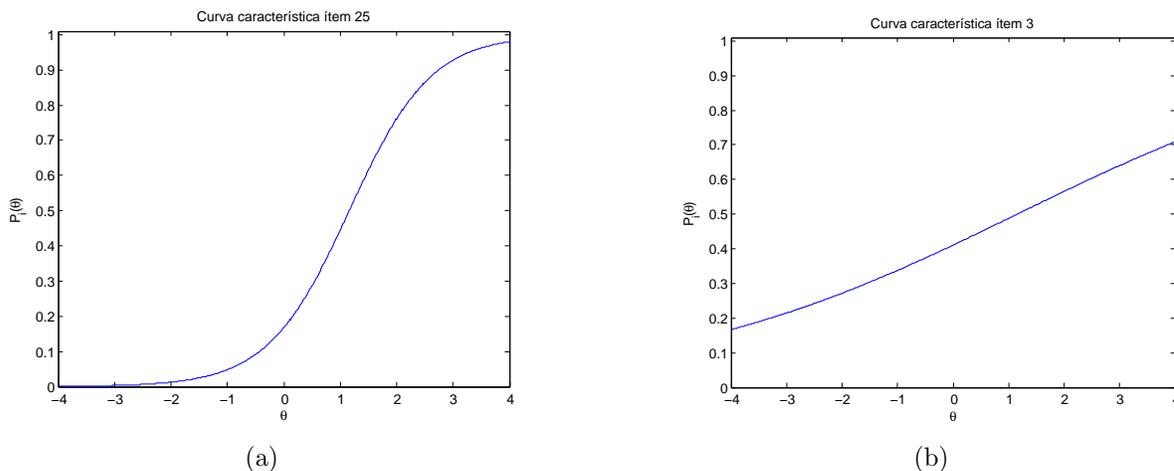


Figura 4.2: Curvas características del ítem para dos ítems con dificultades similares, donde el primero es muy discriminante y el segundo muy poco.

La idea es que los ítems sean lo más discriminantes posibles, de hecho en el caso extremo en que una curva sea plana el ítem es completamente inútil. En el caso que el modelo se estime con la función logística, Baker [2] entrega los siguientes valores referenciales para determinar si las discriminaciones son altas o bajas.

4.3.1. Relación con análisis factorial

La relación entre el modelo de dos parámetros y el análisis factorial es a través de un modelo más general que llamamos modelo de TRI multidimensional. Este modelo considera que las habilidades tienen más de una dimensión, o sea, que forman un vector $\vec{\theta}$ de $m \times 1$. Al igual que los modelos unidimensionales se tiene una versión del modelo con la función ojiva

Nivel de discriminación	Rango de valores de a_i
Ninguno	0
Muy bajo	0,01-0,34
Bajo	0,35-0,64
Moderado	0,65-1,34
Alto	1,35-1,69
Muy alto	>1,7
Perfecto	∞

Tabla 4.2: Niveles referenciales de discriminación dados por Baker en [2]

normal y otra con la función logística. Sólo se presenta la versión con la función de ojiva normal, ya la relación teórica con el análisis factorial sólo se tiene en este caso. Con esto el modelo propone que la probabilidad de responder de forma correcta al ítem i se caracteriza de la siguiente forma:

$$\mathbb{P}_i(\vec{\theta}) = \int_{-\infty}^{\sum_{k=1}^m a_{ik}\theta_k - b_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = p_i(\vec{\theta}) \quad (4.15)$$

Se supone además que se tiene independencia local, es decir que los ítems condicionados a las habilidades son independientes entre sí y que $\vec{\theta}$ se distribuye como una normal multivariada con media cero y matriz de covarianzas la identidad. Luego si $x = (x_1, \dots, x_p)$ es el vector de ítems de dimensión $p \times 1$ y $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_p) \in \{0, 1\}^p$ es una realización del vector x . Se tiene que la probabilidad de $x = \tilde{x}$ condicional a $\vec{\theta}$, gracias al supuesto de independencia local es

$$\mathbb{P}(x = \tilde{x} | \vec{\theta}) = \prod_{i=1}^p (\mathbb{P}_i(\vec{\theta}))^{\tilde{x}_i} (1 - \mathbb{P}_i(\vec{\theta}))^{1 - \tilde{x}_i} \quad (4.16)$$

por último la probabilidad marginal de $x = \tilde{x}$ es:

$$\mathbb{P}(x = \tilde{x}) = \int_{\mathbb{R}^m} \mathbb{P}(x = \tilde{x} | \vec{\theta}) g(\vec{\theta}) d\vec{\theta} \quad (4.17)$$

Donde $g(\vec{\theta})$ es la función de densidad de $\vec{\theta}$, o sea la densidad de una normal multivariada de media 0 y matriz de covarianzas la identidad.

Al modelo dado por (4.15), (4.16) y (4.17) le llamamos Modelo de TRI multidimensional. La relación entre el modelo de análisis factorial y el modelo de TRI multidimensional es que el Modelo de TRI multidimensional definido con la función ojiva normal es equivalente teóricamente a un modelo de análisis factorial sobre la matriz de correlaciones tetracóricas. Este resultado es de Takane y De Leeuw [40]. La demostración consiste en lo siguiente: En el modelo de análisis factorial, si llamamos $x^* = (x_1^*, \dots, x_p^*)$ al vector de variables latentes continuas que definen al vector x de forma que:

$$x_i = \begin{cases} 1 & \text{si } x_i^* > t_i, \\ 0 & \text{si } x_i^* \leq t_i. \end{cases} \quad (4.18)$$

Donde los parámetros t_i son umbrales, resulta que el modelo de análisis factorial sobre la matriz de correlaciones tetracóricas de x , es por definición un modelo de análisis factorial

sobre las variables x^* , lo que resulta en:

$$x^* = \Lambda f + Lu \quad (4.19)$$

Se supone que f es el vector de factores y que se distribuye como una normal multivariada de media 0 y matriz de covarianzas igual a la identidad. Además u se supone que se distribuye como una normal multivariada de media 0 y matriz de covarianzas igual a la identidad y L^2 se asume que es diagonal. Se supone también que x^* tiene media 0. Esto implica que usando (4.19) x^* se distribuye como:

$$x^* \sim N(0, \Lambda\Lambda^t + L^2) \quad (4.20)$$

y x^* condicional a f se distribuye como:

$$x^*|f \sim N(\Lambda f, L^2) \quad (4.21)$$

y usando que las variables en x^* están dicotomizadas en x se tiene:

$$\mathbb{P}(x = \tilde{x}) = \int_R \phi(\bar{x}^*) d\bar{x}^* \quad (4.22)$$

donde $\phi(\bar{x}^*)$ es la función de densidad de x^* , (que se mostró que era normal multivariada). Donde $R = R_1 \times R_2 \times \dots \times R_p$, y cada R_i está definido como:

$$R_i = \begin{cases} (t_i, \infty) & \text{si } \tilde{x}_i = 1 \\ (-\infty, t_i) & \text{si } \tilde{x}_i = 0 \end{cases} \quad (4.23)$$

Vamos a demostrar que el modelo de análisis factorial de x^* implica al modelo de TRI multidimensional, o sea que (4.19),(4.20),(4.21) y (4.22) \Rightarrow (4.15), (4.16) y (4.17). Si denotamos $h(x^*|\bar{f})$ a la densidad condicional de x^* respecto f , y $g(\bar{f})$ a la densidad de f tenemos:

$$\mathbb{P}(x = \tilde{x}) = \int_R \phi(\bar{x}^*) d\bar{x}^* \quad (4.24)$$

$$= \int_R \left(\int_{\mathbb{R}^p} h(\bar{x}^*|\bar{f}) g(\bar{f}) d\bar{f} \right) d\bar{x}^* \quad (4.25)$$

$$= \int_{\mathbb{R}^p} g(\bar{f}) \int_R h(\bar{x}^*|\bar{f}) d\bar{x}^* d\bar{f} \quad (4.26)$$

Si usamos además (4.21) y el supuesto de que L^2 es diagonal, tenemos independencia de las variables x_i^* condicionadas a f , luego:

$$\int_R h(\bar{x}^*|\bar{f}) d\bar{x}^* = \prod_{i=1}^p \int_{R_i} h_i(\bar{x}_i^*|\bar{f}) d\bar{x}_i^* \quad (4.27)$$

$$= \prod_{i=1}^p \left(\int_{t_i}^{\infty} h_i(\bar{x}_i^*|\bar{f}) d\bar{x}_i^* \right)^{\tilde{x}_i} \left(1 - \int_{t_i}^{\infty} h_i(\bar{x}_i^*|\bar{f}) d\bar{x}_i^* \right)^{1-\tilde{x}_i} \quad (4.28)$$

Pero usando de nuevo (4.21) tenemos que:

$$\int_{r_i}^{\infty} h_i(\bar{x}_i^*|\bar{f}) d\bar{x}_i^* = \Phi\left(\frac{\Lambda_{i \bullet} \bar{f} - t_i}{L_{ii}}\right) \quad (4.29)$$

donde L_{ii}^2 es el elemento i -ésimo de la diagonal de L^2 . Ahora si llamamos:

$$f = \bar{\theta} \quad (4.30)$$

$$\frac{\Lambda_{i,\bullet}}{L_{ii}} = (a_{i1}, \dots, a_{im}) \quad (4.31)$$

$$-\frac{t_i}{L_{ii}} = b_i \quad (4.32)$$

Se recupera exactamente (4.15), (4.16) y (4.17), el implica hacia al otro lado es devolver los cálculos ya hechos, por esto que no se detallará. De este resultado podemos concluir la analogía o relación que existe entre los parámetros de ambos modelos:

- Los m factores del análisis factorial y la m habilidades del modelo de IRT general.
- Las discriminaciones (a_{i1}, \dots, a_{im}) con $\Lambda_{i,\bullet}$.
- El supuesto de independencia local es análogo a suponer que L es diagonal y que u se distribuye como una normal multivariada con matriz de covarianzas la identidad. Estos supuestos en ambos modelos lo que dicen es que si se saca el efecto de $\vec{\theta}$ o de los factores comunes f quedan variables independientes entre sí.

Es importante notar que la equivalencia entre modelos es sólo a nivel teórico, y que los distintos procesos de estimación de los métodos pueden llevar a resultados muy diferentes entre sí. Sin embargo como los modelos de TRI multidimensionales en la práctica son muy difíciles de estimar y aún no se masifica su uso [29], este resultado entrega una opción que teóricamente es equivalente a los modelos de TRI multidimensionales y que es más factible de estimar: el análisis factorial sobre las correlaciones tetracóricas.

4.4. Estimación de los parámetros

En el modelo de Rasch los parámetros son indeterminados, ya que si se define $b_i^* = b_i + l$ y $\theta^* = \theta + l$ se cumple que

$$\mathbb{P}_i(\theta|b_i) = \mathbb{P}_i(\theta^*|b_i^*) \quad (4.33)$$

En el modelo de dos parámetros también, ya que si se define $b_i^* = \frac{b_i}{m} + l$, $a_i^* = a_i \cdot m$ y $\theta^* = \frac{\theta}{m} + l$ se tiene que

$$\mathbb{P}_i(\theta|b_i, a_i) = \mathbb{P}_i(\theta^*|b_i^*, a_i^*) \quad (4.34)$$

Es por esto que para estimar los parámetros del modelo se tiene que imponer alguna restricción, lo más común es que se imponga que la media de las dificultades o las habilidades sea igual a 0. También en el modelo de 2PL se tiene que imponer la desviación estándar de las dificultades o las habilidades igual a 1.

La estimación de parámetros en los modelos de TRI usualmente se hace a través de máxima verosimilitud. Sin embargo existen también opciones de estimación bayesiana. En esta sección se explica brevemente como funcionan los métodos de máxima verosimilitud para los modelos de Rash y de 2PL.

Si se tiene un patrón de respuesta para los p ítems $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_p) \in \{0, 1\}^p$. La probabilidad de obtener este patrón, dado una habilidad θ , se escribe gracias al supuesto de independencia local de la siguiente forma

$$\mathbb{P}(x = \tilde{x}|\theta) = \prod_{i=1}^p \mathbb{P}_i(\theta)^{\tilde{x}_i} \cdot (1 - \mathbb{P}_i(\theta))^{1-\tilde{x}_i} \quad (4.35)$$

Si se tienen los patrones de respuestas de los p ítems de N individuos $\tilde{x}^1, \dots, \tilde{x}^N$, los cuales son una muestra aleatoria de una población mayor entonces las distribuciones de x para cada examinado son estadísticamente independientes [26]. Entonces la probabilidad de obtener sus patrones de respuesta es:

$$\mathbb{P}(\tilde{x}^1, \dots, \tilde{x}^N | \theta_1, \dots, \theta_N) = \prod_{j=1}^N \prod_{i=1}^p \mathbb{P}_i(\theta)^{\tilde{x}_i^j} \cdot (1 - \mathbb{P}_i(\theta))^{1-\tilde{x}_i^j} \quad (4.36)$$

A esta probabilidad la llamamos función de verosimilitud, y se denota como:

$$L(\theta_1, \dots, \theta_N, b_1, \dots, b_p, a_1, \dots, a_p) = \mathbb{P}(\tilde{x}^1, \dots, \tilde{x}^N | \theta_1, \dots, \theta_N) \quad (4.37)$$

4.4.1. Estimación de habilidad suponiendo los parámetros de los ítems conocidos

En este caso la función de verosimilitud sólo depende de $\theta_1, \dots, \theta_N$. Luego para obtener los estimadores de estos parámetros basta maximizar $L(\theta_1, \dots, \theta_N, b_1, \dots, b_p, a_1, \dots, a_p) = L(\theta_1, \dots, \theta_N)$. Los estimadores de habilidad obtenidos de esta forma son consistentes, eficientes y se distribuyen asintóticamente de forma normal [18]. En general los parámetros de los ítems no se conocen, pero el método para estimar se utiliza bajo el supuesto de que las estimaciones que se tienen de los parámetros de los ítems son los verdaderos parámetros, este caso se describe más adelante.

4.4.2. Estimación de los parámetros de los ítems y las personas simultáneamente

Este es el caso más común en la TRI y se presentan tres formas de estimación dentro del enfoque de máxima verosimilitud. Las descripciones de estas tres formas de estimación son un resumen de lo expuesto en el libro de Hambleton [18].

1. Estimación conjunta por máxima verosimilitud²: En este tipo de estimación se toma una estimación inicial para los parámetros de los ítems y la función de verosimilitud (ver ecuación (4.36)) se estiman las habilidades utilizando los parámetros de los ítems como conocidos. Luego utilizando las estimaciones de habilidades se estiman de nuevo los parámetros de los ítems, luego se vuelven a estimar las habilidades, etc. Este proceso se repite hasta que las habilidades y los parámetros de los ítems convergen.

Como se está estimando de forma conjunta a los parámetros de los ítems y de habilidades, no se tienen las propiedades de consistencia usuales de la máxima verosimilitud, sin embargo se ha mostrado empíricamente que los estimadores de los parámetros de los ítems y las habilidades son consistentes [39].

2. Estimación de máxima verosimilitud marginal³ : En este tipo de estimación, para eliminar la dependencia de θ , se calcula la distribución marginal de \tilde{x} conociendo θ . Es decir, en vez de maximizarse la ecuación (4.36), se calcula la distribución marginal de \tilde{x} conociendo θ :

$$P(\tilde{x}|b_1, \dots, b_p, a_1, \dots, a_p) = \int_{-\infty}^{\infty} \prod_{i=1}^p \mathbb{P}_i(\theta)^{\tilde{x}_i} \cdot (1 - \mathbb{P}_i(\theta))^{1-\tilde{x}_i} \cdot g(\theta) d\theta = \pi_{\tilde{x}} \quad (4.38)$$

Donde $g(\theta)$ es la función de densidad de θ y $\pi_{\tilde{x}}$ sería la probabilidad de obtener el patrón \tilde{x} en la población. Si $r_{\tilde{x}}$ denota el número de examinados que obtuvieron el patrón de respuesta \tilde{x} , la función de verosimilitud se expresa como:

$$L(b_1, \dots, b_p, a_1, \dots, a_p) = \prod_{\tilde{x} \in \{0,1\}^p} \pi_{\tilde{x}}^{r_{\tilde{x}}} \quad (4.39)$$

Luego los estimadores de máxima verosimilitud marginal se obtienen encontrando los parámetros $b_1, \dots, b_p, a_1, \dots, a_p$ que maximizan a L .

3. Estimación de máxima verosimilitud condicional⁴ : Este tipo de estimación sólo sirve para el modelo de Rasch ya que utiliza el hecho de que en este modelo existen estadísticos suficientes. Lo que hace es expresar en la función de verosimilitud a θ_j en términos del puntaje de la persona j . Con esto la función sólo depende de los parámetros de dificultad y se puede maximizar para encontrarlos. Como el estimador de máxima verosimilitud de las dificultades b_i se puede obtener sin ninguna referencia a las habilidades, este estimador cumple con ser consistente y eficiente [18]. El único problema en este tipo de estimación es que las funciones que quedan al reemplazar θ por una expresión dependiente del puntaje de la persona, son las llamadas funciones elementales simétricas las cuales presentan problemas numéricos en el cálculo de la primera y segunda derivada [18]. Estas derivadas son necesarias para maximizar numéricamente a la verosimilitud.

²En inglés se conoce como *joint maximum likelihood estimation* o *unconditional maximum likelihood*

³En inglés *marginal maximum likelihood estimation*

⁴En inglés *conditional maximum likelihood estimation*

4.5. Función de información y error estándar de medición

La función de información en la TRI tiene relación con la precisión que se tiene para estimar un parámetro. En particular a mayor información se tiene más precisión. Su definición es la de la función de información de Fisher en el caso en que el logaritmo de la verosimilitud es dos veces diferenciable.

Definición 4.8 (Función de información)

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) \quad (4.40)$$

A partir de la expresión de la verosimilitud (ver ecuación (4.36)) , se demuestra que ⁵:

$$I(\theta) = \sum_{i=1}^p \frac{[P_i(\theta)']^2}{P_i(\theta) \cdot (1 - P_i(\theta))} \quad (4.41)$$

Una de la propiedades principales de la función de infomación de test es que la contribución a la información dada por cada ítem es aditiva, esto permite determinar precisamente el aporte de cada ítem a la información que da el test [18]. También permite notar que la calidad de la información depende de la cantidad y calidad de cada ítem que forma al test. Se define la función de información del ítem i como :

Definición 4.9 (Función de información del ítem i)

$$I_i(\theta) = \frac{[P_i(\theta)']^2}{P_i(\theta) \cdot (1 - P_i(\theta))} \quad (4.42)$$

Utilizando las ecuaciones de información (4.41), (4.42) y las ecuaciones logísticas que definen a los modelos (4.9) y (4.14), se tiene que Como la función $f(p) = p(1 - p)$ tiene su

Modelo	Función de información	Expresión
Rasch	Ítem	$I_i(\theta) = P_i(\theta) \cdot (1 - P_i(\theta))$
2PL	Ítem	$I_i(\theta) = a_i^2 P_i(\theta) \cdot (1 - P_i(\theta))$
Rasch	Test	$I(\theta) = \sum_{i=1}^p P_i(\theta) \cdot (1 - P_i(\theta))$
2PL	Test	$I(\theta) = \sum_{i=1}^p a_i^2 P_i(\theta) \cdot (1 - P_i(\theta))$

Tabla 4.3: Funciones de información de ítem y de test en el modelo de Rasch y de 2PL

máximo en $p = 1/2$ y se vió que el valor de θ para el cual $P_i(\theta) = 1/2$ es $\theta = b_i$ en el modelo de Rasch y de 2PL, se tiene que la función de información de cada ítem alcanza su máximo en $\theta = b_i$. Luego un ítem da más información para valores de θ cercanos a b_i . Por otro lado en el modelo de 2PL, por la fórmula de la función de información de cada ítem, es evidente que

⁵El detalle de esta demostración se puede ver en el libro de Hambleton [18] en las páginas 98-99

los ítems más discriminantes son los que entregán más información. El valor máximo para la función de información de test también es importante ya que alrededor de ese valor es donde la prueba mide más precisamente o da más información.

La función de información se relaciona inversamente con el error estándar de medición, el cual se define como:

Definición 4.10 (Error estándar de medición)

$$SEM(\theta) = \frac{1}{I(\theta)} \quad (4.43)$$

Lo más importante del error estándar de medición es que entrega la precisión con que se mide para cada valor de θ , lo que en Teoría Clásica no se tenía.

4.6. Aplicación de la teoría de respuesta al ítem a la prueba CAM

Antes de aplicar los modelos de la TRI a los datos, es necesario verificar cuanto se cumplen los supuestos del modelo [18]. El principal supuesto a verificar es que los datos sean unidimensionales, esto se hizo en el capítulo de Análisis factorial, al aplicar un modelo de análisis factorial sobre la matriz de correlaciones de Pearson y de correlaciones tetracóricas. En el caso de la prueba CAM total, en los gráficos de sedimentación se vió que no era unidimensional (ver Figuras 2.1 y 2.2). Sin embargo el factor que representaba más varianza coincidía con el constructo *conocimiento de alumnos y matemáticas*.

En el caso de la prueba CAM depurada el análisis sobre ambas matrices de correlaciones permitió concluir que esta prueba es unidimensional (Figuras 2.3 y 2.4).

El supuesto de independencia local teóricamente se obtiene de la unidimensionalidad, sin embargo es bueno hacer un análisis de la prueba para determinar si tiene alguna característica que lo rompa. Aunque el diseño de la prueba no tenía ítems que dependieran de otros ítems para responderse, en el análisis factorial de la prueba CAM total, se formaron factores que sólo agruparon ítems similares. Una posible causa para esto es que estos ítems no cumplen con la independencia local. Sin embargo en la prueba CAM depurada estos ítems fueron eliminados y no ocurrió este problema, con esto más la unidimensionalidad mostrada por el análisis factorial se puede asumir que en la prueba CAM depurada se tiene independencia local.

En el modelo de Rasch se considera que todas las discriminaciones son iguales a 1. Se puede tener una noción de cuanto se cumple este supuesto al mirar las discriminaciones entregadas por teoría clásica. En el caso de la prueba CAM total estas son muy dispares ya que van de un rango de 0,42 a -0,01 con una media de 0,21 y una desviación estándar igual a 0,11.

Por otra parte en la prueba CAM depurada las discriminaciones tienen un rango de 0,45 a 0,17, una media de 0,29 y una desviación estándar de 0,07. A pesar de que en este último caso no son completamente uniformes se tiene que por lo menos son todas significativamente positivas y que su desviación estándar es más baja.

Se aplicó el modelo de Rasch y de dos parámetros a la prueba CAM total y a la prueba CAM depurada con el objetivo de analizarlas y describir sus ítems. En la prueba CAM total se encontró una persona con puntaje 0, que fue eliminada antes de los análisis. El que la prueba CAM depurada tuviera menos ítems ocasionó que aparecieron 6 personas con puntaje 0, para los análisis sobre esta prueba estas personas también fueron eliminadas. Se analizaron los estadísticos de ajuste de Rasch en ambas pruebas. Respecto a la función de información, se determinó para que rangos de habilidades la prueba entrega más información.

4.6.1. Estimación del modelo de Rasch

Para la prueba CAM total se presenta la estimación de las dificultades, Infit y Outfit de cada ítem en la Tabla 4.4. Los valores de Infit están todos en un rango que muestra un excelente ajuste de los ítems (todos en el rango $0,8 - 1,2$). En el caso del Outfit todos los valores están en el rango $0,7 - 1,3$ salvo los ítems 13b y 22, sin embargo ambos valores están muy cercanos a 1,3 (ver Tabla 4.4).

Respecto al ajuste de las personas, todas tuvieron un Infit en el rango $0,7 - 1,3$. Respecto al Outfit, 21 personas se superaron el rango $0,7 - 1,3$, de las 21, 3 tuvieron un Outfit mayor que 2.

Ítem	dificultad	Infit MSQ	Outfit MSQ	Ítem	dificultad	Infit MSQ	Outfit MSQ
I13b	2,92	0,97	1,34	I16d	-0,21	1,07	1,07
I1b	1,54	1	1,25	I17b	-0,21	1	0,98
I22	1,47	0,98	1,42	I26c	-0,23	0,95	0,92
I5	1,47	1,05	1,21	I28c	-0,24	0,95	0,93
I9	1,37	1,01	1,22	I26d	-0,26	1,02	1
I8	1,19	0,95	1,23	I3	-0,26	1,03	1,07
I20d	1,05	0,98	0,95	I18a	-0,28	0,97	0,94
I28a	0,95	0,97	0,94	I13a	-0,29	0,95	0,92
I18b	0,86	1,08	1,12	I1a	-0,31	1,09	1,13
I10d	0,81	1,03	1,02	I10b	-0,32	0,93	0,92
I13c	0,73	1	0,96	I1c	-0,37	1,02	1,02
I17d	0,7	0,99	0,94	I13d	-0,38	0,97	0,94
I20b	0,7	0,94	0,86	I17a	-0,38	1,02	1,01
I20c	0,68	0,97	0,91	I24c	-0,4	1,01	1
I25	0,64	0,9	0,91	I28e	-0,5	0,89	0,88
I27	0,62	0,9	0,84	I10e	-0,52	0,94	0,94
I28d	0,62	1,04	1,13	I21	-0,56	0,94	0,94
I20a	0,58	0,97	0,98	I26a	-0,58	0,95	0,93
I15	0,44	1	1	I17c	-0,69	1,02	1
I7	0,44	1	1,11	I6	-0,76	1,1	1,14
I10a	0,37	1,1	1,18	I19a	-0,79	0,91	0,89
I11	0,33	1,08	1,18	I24b	-1,02	0,96	0,97
I18c	0,3	0,99	0,93	I16b	-1,05	1,11	1,14
I26b	0,19	1,09	1,15	I4	-1,12	1,11	1,14
I10c	0,09	1,07	1,1	I14	-1,17	1,08	1,06
I23	0,03	0,97	0,94	I2	-1,38	1,04	1,05
I20e	-0,01	1	0,98	I24a	-1,45	0,9	0,89
I19c	-0,01	0,95	0,93	I19b	-1,52	0,89	0,86
I28b	-0,04	0,94	0,92	I16a	-1,67	0,88	0,85
I24d	-0,18	1,02	1	I18d	-1,72	0,93	0,93
I16c	-0,21	1,05	1,04				

Tabla 4.4: Estadísticos de los ítems: dificultad, Infit y Outfit para la prueba CAM total

El mapa ítems-personas de la prueba CAM total muestra que la prueba fue difícil para la población, ya que aparecen ítems (22, 9, 5, 1b, 20d y 8) cuya dificultad no era similar a la habilidad de ningún examinado. Es decir todos los examinados tienen una probabilidad menor a 1/2 de responder correctamente esas preguntas, incluso los con mayor puntaje.

Por otro lado se tiene que el grupo de examinados con menor habilidad no tienen ítems con una dificultad que permita distinguirlos con máxima información. En todo caso los otros ítems dan información sobre estas personas, sin embargo si se necesitan medidas del constructo con una precisión muy alta para este grupo de examinados se debieran de agregar ítems muy fáciles, con dificultades cercanas a su nivel de habilidad.

Por otro lado las dificultades en el rango en que se distribuyen forman aproximadamente un continuo, es decir no hay espacios en la distribución de las dificultades.

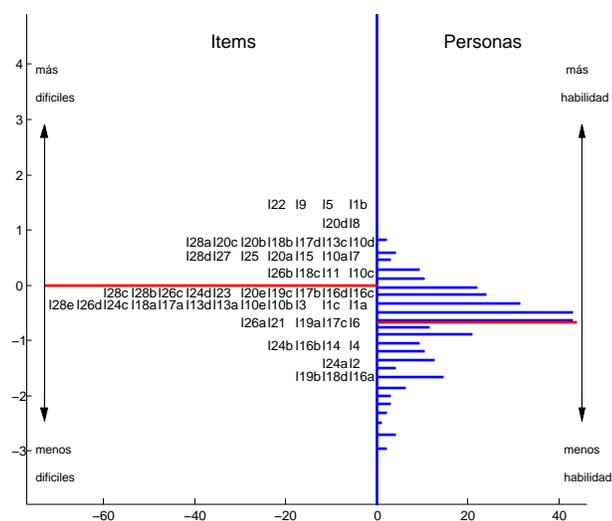


Figura 4.3: Gráfico ítems-personas para la prueba CAM total.

A pesar de que los supuestos de unidimensionalidad, independencia local y discriminaciones iguales no se cumplieron en la prueba CAM total, los estadísticos Infit y Outfit mostraron un buen ajuste. Probablemente se deba a que estos estadísticos fallan frente a violaciones más fuertes de los supuestos. Por la ecuación de estos estadísticos se concluye que a pesar de todo el modelo logró predecir los patrones de respuesta de los ítems y de las personas. Sin embargo por la falla de los supuestos es posible que las estimaciones de los parámetros no sean invariantes.

Para la prueba CAM depurada los estadísticos de ajuste también fueron muy buenos, todos los valores de Infit y Outfit estuvieron dentro de un rango excelente (ver Tabla 4.5).

Respecto al ajuste en las personas, todas tuvieron un Infit entre 0,7 y 1,3 salvo 4 personas que tuvieron un Infit ligeramente sobre 1,3, pero no demasiado (el mayor Infit encontrado fue 1,42). En el Outfit 19 personas superaron el valor 1,3, sin embargo tampoco por mucho, sólo uno fue mayor a 2. También en el Outfit se presentaron 8 personas con un sobre ajuste, es decir con valores menores a 0,7. Sin embargo es más importante determinar un mal ajuste que un sobre ajuste. Como los indicadores de mal ajuste fueron pocos y prácticamente

sólo en el Outfit no se eliminaron personas.

Ítem	dificultad	Infit MSQ	Outfit MSQ	Ítem	dificultad	Infit MSQ	Outfit MSQ
I28a	1,36	1	0,97	I18a	0,06	1,04	1,07
I17d	1,1	1,01	0,98	I10b	0,01	1,01	1,14
I25	1,04	0,87	0,89	I17a	-0,05	1,12	1,18
I27	1,01	0,86	0,76	I24c	-0,07	1,05	1,09
I15	0,83	1,07	1,14	I28e	-0,18	0,87	0,82
I18c	0,67	1,01	0,96	I10e	-0,2	1,04	1,23
I23	0,38	1,03	1,05	I21	-0,25	1	1,03
I19c	0,35	0,94	0,95	I26a	-0,26	0,94	0,91
I28b	0,31	0,97	0,92	I19a	-0,49	0,97	0,92
I24d	0,16	1,06	1,05	I24b	-0,74	1	1,04
I17b	0,13	1,09	1,11	I24a	-1,21	0,9	0,9
I26c	0,11	0,94	0,88	I19b	-1,29	0,92	0,84
I28c	0,09	0,95	0,9	I16a	-1,45	0,96	0,95
I26d	0,08	1,04	1,1	I18d	-1,51	1,01	1,06

Tabla 4.5: Estadísticos de los ítems: dificultad, Infit y Outfit para la prueba CAM depurada

Respecto al gráfico de ítems-personas de la prueba CAM depurada (Figura 4.4) se tiene que para las personas con mayor y menor habilidad no hay ítems que midan con dificultades similares a sus habilidades. Sin embargo hay una cantidad considerable de ítems con dificultades en el rango más alto de dificultad (ítems 27, 25, 17d, 15) y en el más bajo también (ítems 24a, 19b, 18d, 16a), esto es bueno ya que estos grupos de ítems dan información sobre las personas que están en los extremos, luego a pesar de que no hayan ítem con una dificultad exactamente igual a sus habilidades, si hay grupos de ítems en los extremos que informan sobre estas personas.

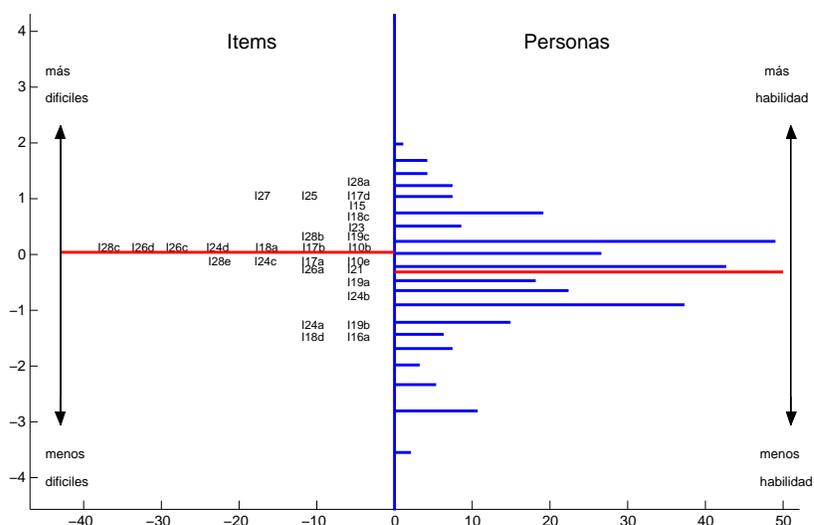


Figura 4.4: Gráfico ítems-personas para la prueba CAM depurada.

4.6.2. Estimación del modelo de 2PL

Se muestra en la Tabla 4.6 las dificultades estimadas por el modelo de dos parámetros para la prueba CAM total. El rango típico de dificultad para los modelos de la TRI usando la función logística está entre -3 y 3 [2]. En el caso de la prueba CAM los ítems más difíciles y más fáciles presentan dificultades del orden de las decenas. Esto definitivamente se sale de lo usual, por eso se inspeccionaron otras características de los ítems y se encontró que los ítems con dificultades extremas tenían discriminaciones nulas o muy bajas (ver Tabla 4.7).

Ítem	Dificultad	Ítem	Dificultad	Ítem	Dificultad	Ítem	Dificultad
I10a	33,28	I28a	2,12	I17b	0,77	I17c	-0,09
I1b	18,51	I15	2,12	I19c	0,72	I19a	-0,13
I9	17,98	I20a	2,07	I28b	0,71	I24b	-0,47
I18b	17,21	I20c	2,04	I26d	0,61	I6	-0,59
I13b	15,48	I17d	2,04	I18a	0,54	I24a	-0,77
I11	5,94	I16d	1,96	I13a	0,53	I19b	-0,83
I1a	4,8	I20b	1,72	I17a	0,51	I16a	-1,03
I26b	4,01	I18c	1,43	I26c	0,49	I14	-1,34
I22	3,99	I16c	1,42	I28c	0,46	I18d	-1,42
I10d	3,57	I20e	1,2	I13d	0,45	I16b	-1,59
I20d	3,49	I25	1,16	I24c	0,4	I2	-1,68
I28d	3,25	I3	1,16	I10b	0,38	I4	-23,87
I13c	3,24	I27	1,11	I10e	0,17	I5	-35,49
I8	3,19	I23	0,94	I28e	0,15		
I10c	2,89	I1c	0,85	I21	0,11		

Tabla 4.6: Dificultades estimadas en el modelo de dos parámetros para la prueba CAM total

Las discriminaciones de los ítems se presentan en la Tabla 4.7. Se puede ver que hay ítems cuya discriminación es nula o muy baja según el criterio dado por Baker (ver Tabla 4.2). Estos ítems son los mismos que tuvieron discriminaciones bajas en Teoría Clásica y correlaciones bajas con el factor de *conocimiento de alumnos y matemáticas*. Esto muestra que los tres métodos estudiados en esta memoria llevan a la misma conclusión respecto cuales ítems tuvieron una discriminación o representación del constructo deficiente.

A pesar de que habían ítems con discriminaciones de teoría clásica muy bajas para la prueba CAM total, igual el ajuste del Infit y Outfit fue bueno para estos ítems. Es más, el ítem 5 tuvo una discriminación ligeramente negativa y sin embargo mostró un buen ajuste según el Infit y el Outfit.

Respecto a las estimaciones de dificultad y discriminación en la prueba CAM depurada, se tiene que las dificultades volvieron a estar en un rango normal (ver Tabla 4.8). También las discriminaciones tomaron valores de bajos a altos según la clasificación de Baker (ver Tabla 4.9). Luego concluimos que todos los ítems en esta prueba tienen buenas características técnicas.

Ítem	Discriminación	Ítem	Discriminación	Ítem	Discriminación	Ítem	Discriminación
I28e	1,61	I24b	0,83	I22	0,53	I10c	0,24
I27	1,46	I28a	0,79	I20e	0,52	I13b	0,22
I25	1,38	I18d	0,79	I15	0,51	I16d	0,2
I19b	1,22	I23	0,75	I17a	0,5	I26b	0,2
I24a	1,19	I18a	0,72	I20d	0,48	I6	0,19
I16a	1,11	I13a	0,69	I7	0,46	I11	0,15
I19a	1,08	I17d	0,68	I17c	0,45	I1b	0,11
I19c	1	I18c	0,68	I2	0,42	I9	0,11
I28c	0,99	I20c	0,66	I13c	0,41	I18b	0,08
I26a	0,97	I26d	0,64	I10d	0,4	I1a	0,06
I10b	0,96	I24c	0,63	I28d	0,38	I10a	0,03
I28b	0,96	I20a	0,6	I14	0,37	I4	0,02
I26c	0,95	I8	0,58	I3	0,31	I5	-0,06
I21	0,91	I13d	0,58	I1c	0,3		
I10e	0,84	I24d	0,57	I16c	0,29		
I20b	0,83	I17b	0,55	I16b	0,24		

Tabla 4.7: Discriminaciones estimadas en el modelo de dos parámetro para la prueba CAM total

Ítem	Dificultad	Ítem	Dificultad	Ítem	Dificultad	Ítem	Dificultad
I17d	2,21	I17b	0,85	I26c	0,41	I26a	0,04
I15	2,2	I24d	0,76	I28c	0,41	I19a	-0,2
I28a	2,18	I19c	0,69	I10b	0,4	I24b	-0,57
I18c	1,49	I28b	0,69	I24c	0,34	I24a	-0,88
I25	1,13	I18a	0,6	I10e	0,14	I19b	-0,95
I23	1,05	I26d	0,54	I28e	0,1	I16a	-1,37
I27	1,03	I17a	0,53	I21	0,06	I18d	-1,92

Tabla 4.8: Dificultades estimadas en el modelo de dos parámetros para la prueba CAM depurada

4.6.3. Análisis de la función de información

En esta parte se muestra el gráfico de la función de información para la prueba CAM total y depurada en el modelo de Rasch y de 2 parámetros. Cada gráfico de la función de información va acompañado de un histograma de la habilidad en la población. Esto se hizo para ver si donde las pruebas dan máxima información coincide con el rango de habilidad donde hay más frecuencia de personas. Un problema que tuvo Hill [22] en su intento por medir el *conocimiento de alumnos y matemáticas* fue que la habilidad para la cual la función de información era máxima o $\text{argmax}(I(\theta))$, estaba a una distancia aproximada de 2 desviaciones estándares por debajo de la media del grupo. Luego la medición en habilidades mayores a la media tenía mucho error (ver [22]).

En el caso de la prueba CAM total y depurada el máximo de la función de información en ambos modelos estuvo a una distancia menor de una desviación estándar de la media (ver

Ítem	Discriminación	Ítem	Discriminación	Ítem	Discriminación	Ítem	Discriminación
I28e	1,65	I26a	1,04	I28a	0,76	I17d	0,6
I27	1,64	I19a	1	I10b	0,75	I18d	0,6
I25	1,42	I19c	1	I26d	0,67	I24d	0,56
I19b	1,13	I28b	0,93	I10e	0,63	I18a	0,55
I24a	1,12	I16a	0,84	I18c	0,62	I15	0,47
I26c	1,11	I21	0,8	I24c	0,62	I17b	0,45
I28c	1,04	I24b	0,78	I23	0,62	I17a	0,41

Tabla 4.9: Discriminaciones estimadas en el modelo de dos parámetros para la prueba CAM depurada

Tablas 4.10 y 4.11).

Lo que se puede ver con más detalle en las Figuras 4.5 y 4.6 al comparar las curvas de información y los histogramas que muestran que donde se tiene mayor información sobre las habilidades es aproximadamente donde hay más personas.

Prueba	CAM total	CAM depurada
Media de $\hat{\theta}$ en la muestra	-0,67	-0,33
Desviación estándar de $\hat{\theta}$ en la muestra	0,69	0,96
Argmax($I(\theta)$)	-0,03	0,04
Max($I(\theta)$)	13,18	6,24

Tabla 4.10: Estadísticos descriptivos de la habilidad y la información en el modelo de Rasch para la prueba CAM total y depurada

Prueba	CAM Total	CAM depurada
Media de $\hat{\theta}$ en la muestra	0	0,01
Desviación estándar de θ en la muestra	0,9	0,87
Argmax($I(\theta)$)	0,66	0,38
Max($I(\theta)$)	12,71	6,18

Tabla 4.11: Estadísticos descriptivos de la habilidad y la información en el modelo de dos parámetros para la prueba CAM total y depurada

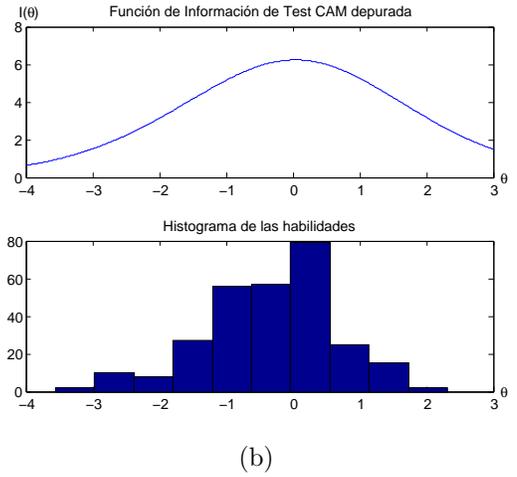
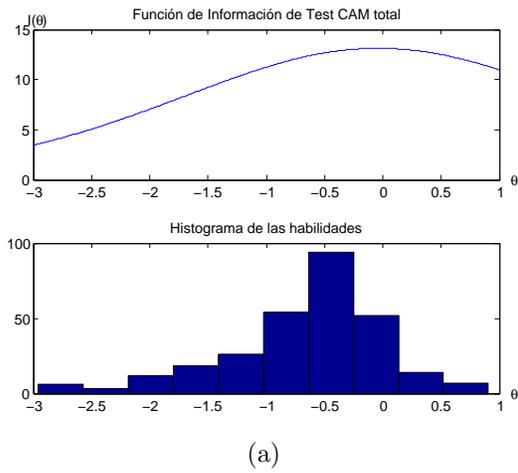


Figura 4.5: Funciones de información e histogramas de la habilidad para la prueba CAM total y depurada en el modelo de Rasch.

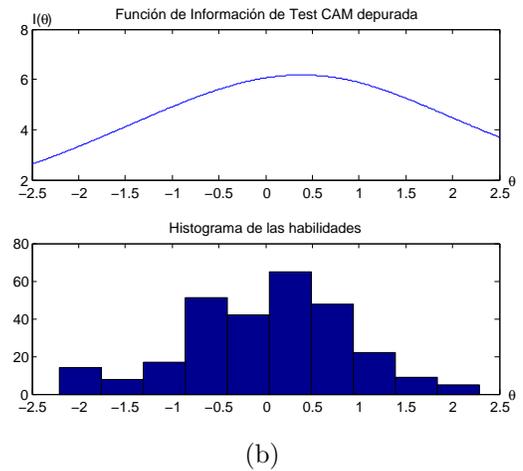
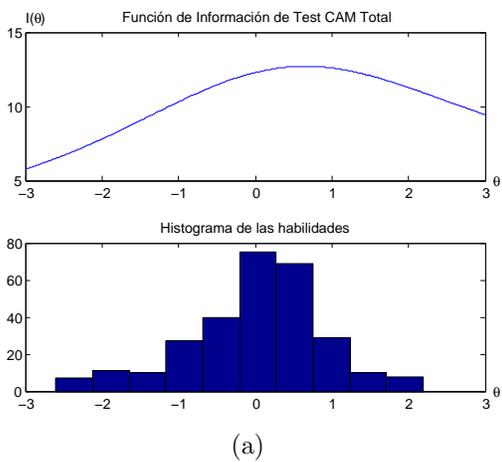


Figura 4.6: Funciones de información e histogramas de la habilidad para la prueba CAM total y depurada en el modelo de dos parámetros.

Capítulo 5

Conclusiones

En esta memoria se estudiaron modelos matemáticos para analizar pruebas y se aplicaron a una prueba nueva sobre *conocimiento de alumnos y matemáticas*. Se presentan primero conclusiones respecto a los métodos y luego sobre los resultados sobre la prueba CAM.

Se estudiaron tres métodos para analizar pruebas, los cuales tienen importantes relaciones teóricas que permiten interpretar de manera más íntegra los resultados que entregan. Las conclusiones específicas sobre cada método se presentan a continuación:

1. Análisis Factorial: De los métodos estudiados, este es el único que permite explorar las dimensiones que está midiendo un conjunto de variables. Esto es muy importante en psicometría y medición educacional ya que permite dilucidar realmente que es lo que está midiendo un instrumento. Gracias a esta propiedad también es la mejor opción para verificar si se cumple el supuesto de unidimensionalidad de TRI. A pesar de que también existen modelos en la TRI para modelar datos multidimensionales, aún no tienen un desarrollo suficiente en estimación ni en softwares. Por otro lado el resultado de Takane y De Leeuw [40] muestra que el análisis factorial sobre la matriz de correlaciones tetracóricas es equivalente al modelo multidimensional de la TRI, por lo que más aún el análisis factorial es una buena alternativa.
2. Teoría Clásica de Test: Esta teoría es una herramienta exploratoria muy potente, que permite evaluar y describir de manera general a una prueba y permite identificar de manera más fácil a los ítems que no están de acuerdo con la escala. Los parámetros son más fáciles de interpretar ya que son principalmente correlaciones y porcentajes, luego es más accesible. Por otro lado es la teoría que menos datos necesita (en comparación con el análisis factorial o la TRI), luego si se tienen pocos datos, por ejemplo 50 personas, es mejor utilizar la teoría clásica. Por último el concepto de confiabilidad originado de esta teoría sigue siendo muy utilizado y es un requisito reportarla al usar cualquier instrumento de medición, sin embargo se suele reportar sólo con el estadístico α de Cronbach. En el desarrollo de este trabajo se encontraron diversos índices que miden mejor o tienen otras ventajas para medir la confiabilidad que el coeficiente α .
3. Teoría de Respuesta al Ítem: Esta teoría es un modelo cuya fortaleza principal es que los estadísticos son invariantes, este es un aporte fundamental para el desarrollo de mediciones. Además de tener muchas propiedades matemáticas excelentes, como

por ejemplo que se puede caracterizar la información que aporta cada ítem y el test completo. También en el modelo de Rasch se pueden comparar a los ítems y a las personas y en el de dos parámetros se puede estimar la discriminación de cada ítem. La principal conclusión respecto a esta teoría es que tiene propiedades y potencialidades fundamentales para medir.

La aplicación de los métodos a la prueba CAM mostró que es un instrumento multidimensional, teniendo sin embargo un factor principal que representa al constructo *conocimiento de alumnos y matemáticas* y otros factores que representan a grupos de ítems que se comportan de forma similar. También se vió que esta prueba tiene una confiabilidad buena, y que la habilidad para la cual da más información está en los rangos de habilidades donde hay mayor frecuencia de personas. Sin embargo esta prueba tiene ítems con baja discriminación que aportan poco a la confiabilidad. Estos mismos ítems tienen bajas correlaciones con el factor que representa al constructo.

Esto motivó a formar a la prueba CAM depurada a partir de una selección de los ítems. Esta se realizó después de identificar al factor que representaba al *conocimiento de alumnos y matemáticas* y el criterio fue escoger a los ítems que correlacionan más alto con este factor. Esta forma de seleccionar a los ítems reporta las siguientes ventajas teóricas: Son los ítems que más representan al constructo y gracias al resultado de Takane [40], los ítems con correlaciones más altas con el factor 1 son también los más discriminantes. También tener una escala más uniforme mejora la confiabilidad. Estas ventajas teóricas se vieron reflejadas en los resultados. Por otra parte, la prueba CAM depurada mostró las siguientes características

1. Representa al constructo *conocimiento de alumnos y matemáticas*.
2. Es unidimensional.
3. Tiene una confiabilidad equivalente a la prueba total a pesar de tener menos ítems.
4. Las discriminaciones de todos los ítems fueron significativamente positivas en el caso de las entregadas por teoría clásica y entre bajas y altas según la TRI.
5. Mide algo distinto a la prueba LMT, es decir no mide sólo habilidad matemática.
6. Los valores de θ para los cuales la prueba da máxima información están cercanos a la media de la población (a menos de una desviación estándar de distancia). Esto para el modelo de Rasch y de dos parámetros.

Por otra parte las relaciones teóricas de los métodos se verificaron en los resultados estimados. Esto significa que hay una coherencia en los datos. Sin embargo no significa que se tenga que usar sólo un método ya que como se vió cada uno entrega aportes particulares.

Respecto al trabajo futuro, sobre la investigación de métodos para analizar pruebas se podría indagar respecto a la Teoría de la Generabilidad que se puede ver como una extensión de la Teoría Clásica. Respecto a la investigación sobre la prueba CAM, se tienen las siguientes propuestas de trabajo a futuro:

1. Bancos de Ítems: La propiedad de invarianza de la TRI y la capacidad que tiene para poder relacionar el puntaje de dos pruebas permite hacer bancos de ítems. Luego los ítems validados de la prueba CAM depurada pueden servir como base para calibrar a otros ítems.

2. Funcionamiento diferencial: La teoría de respuesta al ítem permite ver si un ítem funciona diferente para dos poblaciones. Esto es más sutil que simplemente ver para que población fue más fácil o difícil, ya que se comparan las curvas características de ítems. Una posible investigación es ver si existe un funcionamiento diferencial entre personas con formación en pedagogía y personas con formación en matemáticas.
3. Relacionar la prueba con otras variables: Se puede utilizar a la prueba CAM para comparar al constructo *conocimiento de alumnos y matemáticas* con otras variables como experiencia, formación, o calidad matemática de la instrucción.

Bibliografía

- [1] AERA/APA/NCME. *Standards for educational and psychological testing*. American Educational Research Association, Washington DC, 1999.
- [2] Francis B. Baker. *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, 2nd edition, 2001.
- [3] Deborah L. Ball, Mark Thames, and Geoffrey Phelps. Content knowledge for teaching. What makes it special? *Journal of Teacher Education*, 59(5):389–407, 2008.
- [4] Alexander Basilevsky. *Statistical factor analysis and related methods*. John Wiley & Sons, 1994.
- [5] J. Baumert, M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, S. Krauss, M. Neubrand, and Y. Tsai. Teachers Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Education Research Journal*, 47, 2010.
- [6] Peter Bentler. Alpha dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1):137–143, 2009.
- [7] Trevor G. Bond and Christine M. Fox. *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates, 2007.
- [8] Raymond B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.
- [9] Lee Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951.
- [10] Robert Cudeck and Robert MacCallum, editors. *Factor analysis at 100 Historical developments and future directions*. Lawrence Erlbaum Associates, 2007.
- [11] Edward E. Cureton and Ralph B. D’Agostino. *Factor analysis: an applied approach*. Lawrence Erlbaum Associates, 1983.
- [12] John Fox. Structural equation modeling with the sem package in R. *Structural Equation Modeling*, 13(3):465–486, 2006.

- [13] Richard Gorsuch. *Factor analysis*. Saunders, 1974.
- [14] James Graham. Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement*, 66(6):930–944, 2006.
- [15] Harold Gulliksen. *Theory of mental test*. John Wiley & Sons, New York, 1950.
- [16] Louis Guttman. A basis for analyzing test-retest reliability. *Psychometrika*, 10(2):255–282, 1945.
- [17] Kenneth D. C. Haley. Estimation of the dosage mortality relationship when the dose is subject to error, 1952.
- [18] Ronald K. Hambleton and Hariharan Swainathan. *Item response theory*. Kluwer Nijhoff Publishing, 1985.
- [19] Harry H. Harman. *Modern factor analysis*. The University of Chicago Press, Chicago, 1960.
- [20] Heather Hill, Brian Rowan, and Deborah Ball. Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2):371–406, 2005.
- [21] Heather Hill, Stephen Schilling, and Deborah Ball. Developing measures of teachers’ mathematical knowledge for teaching. *The Elementary School Journal*, 105(1):11–30, 2004.
- [22] Heather C Hill, Deborah Loewenberg Ball, and Stephen G Schilling. Unpacking pedagogical content knowledge : Conceptualizing and measuring teachers topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4):372–400, 2008.
- [23] I. T. Jolliffe. *Principal component analysis*. Springer, 2nd edition, 2002.
- [24] Karl G. Joreskog. Statistical analysis of sets of congeneric test. *Psychometrika*, 36(2):109–133, 1971.
- [25] Karl G. Joreskog. *Factor analysis by MINRES*. 2003.
- [26] Frederic Lord. *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1980.
- [27] Frederic Lord and Melvin Novick. *Statistical theories of mental test scores*. Addison-Wesley, 1968.
- [28] Roderick McDonald. *Test theory*. Lawrence Erlbaum Associates, 1999.
- [29] José Muñiz. *Introducción a la teoría de respuesta a los ítems*. Pirámide, Madrid, 1997.
- [30] Stanley A. Mulaik. *The foundations of factor analysis*. McGraw-Hill, 1972.

- [31] Stanley A. Mulaik. *Linear causal modeling with structural equations*. Chapman & Hall, 2009.
- [32] Daniel Peña. *Análisis de datos multivariantes*. McGraw-Hill, 2003.
- [33] William Revelle. *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois, 2012. R package version 1.2.1.
- [34] Richard A. Reymont and Karl G. Joreskog. *Applied factor analysis in the natural sciences*. Cambridge University Press, 3 edition, 1996.
- [35] Karin Schermelleh-engel, Helfried Moosbrugger, and Hans Müller. Evaluating the fit of structural equation models : tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2):23–74, 2003.
- [36] Lee Shulman. Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2):4–14, 1986.
- [37] Kenneth Shultz and David Whitney. *Measurement theory in action*. SAGE, 2005.
- [38] Gregor Socan. Assesment of reliability when test items are not essentially τ -equivalent. *Developments in Survey Methodology*, pages 23–35, 2000.
- [39] Hariharan Swaminathan and Janice Gifford. Estimation of parameters in the three-parameter latent trait model. Technical report, New York, 1983.
- [40] Yoshio Takane and Jan De Leeuw. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3):393–408, 1987.
- [41] Wei Tang and Ying Cui. A simulation study for comparing three lower bounds to reliability, 2012.
- [42] Jos ten Berge and Frits Zegers. A series of lower bounds to the reliability of a test. *Psychometrika*, 43(4):575–579, 1978.
- [43] Barry Thompson, Samuel Green, and Yanyun Yang. Assessment of the maximal split-half coefficient to estimate reliability. *Educational and Psychological Measurement*, 70(2):232–251, 2010.
- [44] Bruce Thompson. *Exploratory and confirmatory factor analysis*. American Psychological Association, 2004.
- [45] Leonor Varas, Nancy Lacourly, Alejandro López, and Valentina Giaconi. Evaluación del conocimiento pedagógico del contenido para enseñar matemáticas elementales. *Enseñanza de las Ciencias*, 2012.
- [46] Benjamin Wright and Mark Stone. *Measurement essentials*. Wide Range Inc, 2nd edition, 1999.
- [47] Richard Zinbarg, William Revelle, Iftah Yovel, and Wen Li. Cronbach α , Revelles’s β and

McDonald ω_h : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1):123–133, 2005.

Apéndice A

Resultados análisis factorial

A.1. Resultados del análisis de factores principales sobre la matriz de correlaciones de Pearson

Tabla A.1: Valores propios del ACP sobre la matriz de correlaciones de Pearson.

factor	Valor propio	factor	Valor propio
f1	5,33	f31	0,81
f2	3,26	f32	0,77
f3	2,55	f33	0,74
f4	2,28	f34	0,7
f5	2,06	f35	0,68
f6	1,84	f36	0,67
f7	1,79	f37	0,66
f8	1,7	f38	0,65
f9	1,66	f39	0,59
f10	1,61	f40	0,56
f11	1,57	f41	0,55
f12	1,45	f42	0,54
f13	1,43	f43	0,51
f14	1,33	f44	0,49
f15	1,28	f45	0,48
f16	1,24	f46	0,47
f17	1,19	f47	0,45
f18	1,19	f48	0,44
f19	1,15	f49	0,42
f20	1,13	f50	0,38
f21	1,03	f51	0,37
f22	1,02	f52	0,36
f23	1	f53	0,35
f24	0,98	f54	0,33

Continúa en la página siguiente

Tabla A.1 – continúa de la página anterior

factor	Valor propio	factor	Valor propio
f25	0,93	f55	0,32
f26	0,91	f56	0,3
f27	0,88	f57	0,29
f28	0,85	f58	0,26
f29	0,83	f59	0,25
f30	0,82	f60	0,2
–	–	f61	0,1

Tabla A.2: Columnas ordenadas de la matriz de estructura producida por el ACP sobre la matriz de correlaciones de Pearson

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
28e	0,55	13d	0,49	20c	0,64	22	0,31	10d	0,44
19b	0,49	13a	0,49	20a	0,54	27	0,27	10c	0,38
24a	0,48	1b	0,47	20b	0,53	7	0,26	4	0,36
16a	0,47	1a	0,45	20d	0,52	10e	0,25	10a	0,36
19a	0,46	1c	0,43	13b	0,23	25	0,24	16c	0,3
27	0,45	20a	0,42	16a	0,19	20a	0,22	1c	0,29
25	0,44	4	0,4	18b	0,18	10b	0,19	1a	0,26
10b	0,44	20c	0,4	18d	0,17	15	0,18	18d	0,26
26a	0,42	20d	0,37	18a	0,15	9	0,18	26b	0,23
21	0,41	13c	0,35	10a	0,13	1c	0,18	20e	0,22
19c	0,41	20b	0,33	17c	0,13	1a	0,17	5	0,2
10e	0,41	3	0,3	23	0,12	1b	0,17	18a	0,2
28c	0,41	9	0,28	18c	0,11	26a	0,16	16a	0,2
28b	0,4	13b	0,27	10c	0,11	20b	0,15	17c	0,19
26c	0,39	10b	0,24	1b	0,11	20c	0,15	10e	0,18
24b	0,39	2	0,15	24a	0,11	26d	0,13	17d	0,16
18d	0,36	11	0,15	16d	0,1	10d	0,13	3	0,15
13a	0,36	10e	0,14	22	0,1	28e	0,12	1b	0,14
18a	0,35	6	0,11	14	0,09	18c	0,11	9	0,12
20b	0,34	8	0,11	26b	0,09	19c	0,09	17b	0,12
23	0,34	16c	0,07	20e	0,09	3	0,08	18c	0,11
13d	0,31	19a	0,06	28d	0,09	24c	0,08	27	0,1
24c	0,29	19b	0,02	24d	0,08	2	0,08	16d	0,09
26d	0,29	5	0,02	17d	0,07	8	0,08	21	0,06
18c	0,29	7	0,01	7	0,07	6	0,07	13b	0,06
20c	0,29	14	0,01	8	0,06	20d	0,06	26d	0,05
28a	0,28	16d	0,01	24c	0,05	11	0,04	23	0,05
17b	0,28	21	0	16b	0,04	4	0,03	22	0,03
20a	0,27	18a	-0,02	26d	0,02	18a	0,03	24c	0,03
24d	0,27	17c	-0,03	17b	0,02	17d	0,03	18b	0,02

Continúa en la página siguiente

Tabla A.2 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
17d	0,27	15	-0,04	24b	0	16d	0,02	24d	0,01
17a	0,25	16b	-0,04	28a	-0,01	5	0,01	24b	0
20e	0,25	25	-0,04	28c	-0,04	26c	0,01	24a	-0,01
17c	0,24	19c	-0,05	21	-0,04	18b	-0,01	26c	-0,01
2	0,24	16a	-0,05	28b	-0,05	28c	-0,02	10b	-0,02
8	0,24	17a	-0,06	17a	-0,06	23	-0,03	20b	-0,03
7	0,23	17b	-0,07	13c	-0,07	13b	-0,03	28b	-0,03
15	0,23	10d	-0,09	2	-0,07	21	-0,04	19c	-0,03
20d	0,2	17d	-0,1	15	-0,07	19a	-0,04	20a	-0,04
14	0,2	22	-0,1	16c	-0,08	17b	-0,05	15	-0,05
13c	0,2	27	-0,11	28e	-0,09	28b	-0,05	19a	-0,05
3	0,19	28e	-0,12	10d	-0,11	24d	-0,06	20c	-0,05
1c	0,18	23	-0,13	13a	-0,11	20e	-0,07	28d	-0,06
22	0,18	28d	-0,14	26a	-0,12	28a	-0,07	26a	-0,07
10d	0,16	24a	-0,14	5	-0,14	19b	-0,09	7	-0,07
16c	0,15	28b	-0,14	19b	-0,16	16b	-0,12	13d	-0,08
28d	0,15	20e	-0,18	13d	-0,17	26b	-0,13	2	-0,09
16b	0,15	18d	-0,19	26c	-0,17	24b	-0,13	19b	-0,09
6	0,13	18c	-0,19	25	-0,17	24a	-0,14	13c	-0,09
10c	0,12	28a	-0,19	10e	-0,18	14	-0,15	13a	-0,11
16d	0,11	26a	-0,2	19c	-0,18	16c	-0,15	20d	-0,13
26b	0,09	18b	-0,2	9	-0,19	17a	-0,16	28c	-0,14
11	0,08	26d	-0,2	1c	-0,19	18d	-0,16	6	-0,18
9	0,07	24c	-0,21	27	-0,22	17c	-0,17	17a	-0,21
1b	0,07	24b	-0,21	10b	-0,23	16a	-0,17	8	-0,24
1a	0,06	10c	-0,22	1a	-0,24	28d	-0,18	14	-0,25
13b	0,06	26b	-0,22	19a	-0,24	10c	-0,29	28e	-0,26
4	0,04	10a	-0,24	6	-0,27	10a	-0,32	28a	-0,26
18b	0,02	24d	-0,28	11	-0,3	13c	-0,57	25	-0,28
10a	0,01	28c	-0,28	3	-0,3	13a	-0,59	16b	-0,28
5	-0,03	26c	-0,33	4	-0,38	13d	-0,6	11	-0,34

A.2. Resultados del ACP sobre la matriz de correlaciones tetracóricas

Tabla A.3: Valores propios del ACP sobre la matriz de correlaciones tetracóricas

factor	Valor propio	factor	Valor propio
f1	8,17	f32	0,56
f2	4,95	f33	0,48
f3	3,58	f34	0,44
f4	2,97	f35	0,4
f5	2,82	f36	0,38
f6	2,62	f37	0,37
f7	2,27	f38	0,32
f8	2,18	f39	0,25
f9	2,13	f40	0,24
f10	2	f41	0,16
f11	1,93	f42	0,14
f12	1,74	f43	0,1
f13	1,71	f44	0,08
f14	1,55	f45	0,07
f15	1,4	f46	0,01
f16	1,36	f47	0
f17	1,3	f48	0
f18	1,3	f49	0
f19	1,18	f50	0
f20	1,15	f51	0
f21	1	f52	0
f22	0,98	f53	0
f23	0,95	f54	0
f24	0,92	f55	0
f25	0,82	f56	0
f26	0,8	f57	0
f27	0,75	f58	0
f28	0,68	f59	0
f29	0,65	f60	0
f30	0,6	f61	0
f31	0,57		0

Tabla A.4: Columnas ordenadas de la matriz de estructura producida por el ACP sobre la matriz de correlaciones tetracóricas

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
28e	0,65	1b	0,69	4	0,51	13c	0,69	10d	0,48	5	0,56
19b	0,61	13b	0,56	5	0,43	13d	0,63	5	0,4	17a	0,4
16a	0,6	1a	0,53	9	0,41	13a	0,59	10a	0,39	20d	0,37
24a	0,6	13a	0,52	11	0,4	10a	0,4	4	0,38	28d	0,32
25	0,57	1c	0,52	3	0,39	10c	0,37	16c	0,38	28b	0,32
27	0,57	20a	0,51	1a	0,34	28d	0,23	10c	0,36	15	0,26
19a	0,54	13d	0,5	6	0,33	16c	0,22	16a	0,3	20a	0,25
10b	0,52	20d	0,48	10b	0,32	17a	0,18	18d	0,3	18c	0,24
26a	0,5	20c	0,48	19a	0,31	16a	0,18	1a	0,29	16b	0,24
19c	0,5	4	0,44	27	0,28	17c	0,17	17c	0,26	9	0,24
10e	0,49	20b	0,42	10e	0,24	19b	0,17	1b	0,24	14	0,24
21	0,49	13c	0,4	19b	0,24	19a	0,16	1c	0,24	19b	0,23
28c	0,49	9	0,38	1c	0,23	24a	0,15	18a	0,23	17c	0,21
28b	0,48	3	0,33	25	0,22	24b	0,14	20e	0,22	16a	0,21
18d	0,48	10b	0,26	26c	0,19	18d	0,14	18c	0,21	20b	0,19
24b	0,47	2	0,21	15	0,19	14	0,13	9	0,2	18b	0,17
26c	0,47	8	0,18	26a	0,18	26b	0,12	26b	0,14	1b	0,15
20b	0,45	10e	0,13	19c	0,18	4	0,1	10e	0,14	16d	0,15
18a	0,43	11	0,13	10d	0,14	16b	0,1	17d	0,14	20c	0,13
23	0,42	6	0,1	28b	0,13	28a	0,1	3	0,12	26a	0,12
13a	0,4	16c	0,07	17a	0,11	28b	0,09	16d	0,11	11	0,12
28a	0,38	19a	0,06	16c	0,09	21	0,08	17b	0,08	7	0,1
20c	0,37	16d	0,03	13d	0,06	17b	0,07	18b	0,08	19a	0,09
17d	0,37	19b	0,02	28e	0,05	20e	0,06	27	0,07	28a	0,06
26d	0,36	14	0,02	21	0,02	5	0,06	23	0,07	16c	0,06
24c	0,36	7	0,01	18c	0,02	13b	0,03	20b	0,07	24c	0,06
18c	0,36	21	0,01	28c	0,01	24d	0,03	26d	0,07	24a	0,06
17b	0,35	18a	0,01	2	0,01	26c	0,03	20a	0,07	18d	0,02
20a	0,35	17c	-0,03	7	0,01	3	0,02	21	0,06	10a	0
13d	0,35	16a	-0,03	24c	-0,01	28c	0,01	22	0,06	1a	-0,03
8	0,34	19c	-0,04	1b	-0,02	23	0	24c	0,05	23	-0,04
24d	0,34	25	-0,04	26d	-0,02	6	0	28b	0,04	18a	-0,04
20e	0,32	17b	-0,05	28a	-0,02	11	-0,01	20c	0,02	3	-0,06
17a	0,3	15	-0,06	13a	-0,03	17d	-0,01	24a	0,02	13d	-0,06
7	0,3	16b	-0,07	16b	-0,03	19c	-0,03	28d	0	6	-0,07
2	0,3	17d	-0,09	24b	-0,05	16d	-0,04	15	0	26c	-0,07
17c	0,3	17a	-0,1	13c	-0,06	9	-0,07	10b	-0,01	4	-0,07
15	0,28	5	-0,1	17b	-0,07	18a	-0,07	19a	-0,03	8	-0,07
22	0,28	10d	-0,11	28d	-0,07	2	-0,07	26c	-0,04	10b	-0,08
20d	0,27	28e	-0,11	16d	-0,09	18b	-0,08	19c	-0,04	21	-0,08
14	0,25	27	-0,12	14	-0,11	1a	-0,09	24b	-0,04	28e	-0,08
13c	0,24	22	-0,13	8	-0,11	10d	-0,1	20d	-0,04	19c	-0,09

Continúa en la página siguiente

Tabla A.4 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
3	0,23	24a	-0,13	22	-0,12	1c	-0,11	26a	-0,05	20e	-0,09
1c	0,22	23	-0,13	17c	-0,13	28e	-0,12	7	-0,05	13a	-0,09
10d	0,22	28d	-0,18	17d	-0,14	24c	-0,13	19b	-0,06	26d	-0,09
28d	0,2	28b	-0,18	20e	-0,15	26d	-0,13	24d	-0,07	28c	-0,1
16b	0,19	20e	-0,2	24a	-0,15	20d	-0,14	13b	-0,08	22	-0,1
16c	0,18	26a	-0,22	23	-0,16	8	-0,15	17a	-0,15	24b	-0,11
6	0,15	26d	-0,22	18a	-0,18	10b	-0,16	13d	-0,15	25	-0,11
16d	0,14	18d	-0,24	26b	-0,19	26a	-0,16	2	-0,18	17b	-0,11
10c	0,14	24b	-0,25	10a	-0,2	18c	-0,19	13a	-0,19	13c	-0,12
26b	0,12	28a	-0,27	18b	-0,2	1b	-0,19	13c	-0,2	10c	-0,22
11	0,1	10c	-0,27	10c	-0,2	15	-0,19	28c	-0,2	17d	-0,25
9	0,09	24c	-0,27	24d	-0,21	20b	-0,23	6	-0,26	24d	-0,26
1b	0,09	10a	-0,28	16a	-0,22	20c	-0,24	16b	-0,27	1c	-0,28
13b	0,09	18b	-0,28	18d	-0,23	27	-0,24	14	-0,29	2	-0,29
1a	0,07	26b	-0,29	20a	-0,43	25	-0,24	28e	-0,3	27	-0,29
4	0,04	24d	-0,29	20b	-0,47	10e	-0,25	28a	-0,35	10e	-0,31
18b	0,02	18c	-0,32	13b	-0,48	20a	-0,32	25	-0,35	26b	-0,31
10a	0,02	28c	-0,35	20d	-0,49	7	-0,36	8	-0,37	10d	-0,38
5	-0,07	26c	-0,4	20c	-0,59	22	-0,49	11	-0,4	13b	-0,45

A.3. Resultados del análisis análisis factorial estimado con máxima verosimilitud sobre la matriz de correlaciones de Pearson

Tabla A.5: Valores propios del análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson

factor	Valor propio	factor	Valor propio
f1	4,58	f32	-0,05
f2	2,66	f33	-0,05
f3	1,99	f34	-0,08
f4	1,69	f35	-0,1
f5	1,25	f36	-0,12
f6	1,01	f37	-0,15
f7	0,92	f38	-0,19
f8	0,84	f39	-0,21
f9	0,8	f40	-0,22
f10	0,75	f41	-0,24
f11	0,7	f42	-0,25
f12	0,61	f43	-0,27
f13	0,59	f44	-0,3
f14	0,47	f45	-0,31
f15	0,45	f46	-0,32
f16	0,41	f47	-0,33
f17	0,37	f48	-0,35
f18	0,35	f49	-0,37
f19	0,3	f50	-0,4
f20	0,27	f51	-0,42
f21	0,22	f52	-0,45
f22	0,19	f53	-0,46
f23	0,16	f54	-0,48
f24	0,15	f55	-0,49
f25	0,09	f56	-0,5
f26	0,09	f57	-0,52
f27	0,06	f58	-0,54
f28	0,04	f59	-0,55
f29	0,02	f60	-0,59
f30	0,01	f61	-0,63
f31	-0,02		

Tabla A.6: Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
20c	0,65	13a	0,92	26c	0,44	4	0,48	10a	0,39
20a	0,56	13d	0,9	28c	0,37	1c	0,47	10c	0,32
20b	0,55	13c	0,55	28e	0,37	1a	0,47	18d	0,3
20d	0,4	10b	0,26	26a	0,35	3	0,37	4	0,27
28e	0,37	19b	0,25	27	0,34	9	0,35	16c	0,25
16a	0,33	19a	0,22	24b	0,31	10b	0,32	16a	0,24
24a	0,31	20b	0,22	28b	0,3	1b	0,29	10d	0,23
27	0,29	20c	0,22	25	0,3	10e	0,28	17c	0,2
25	0,29	16a	0,2	19b	0,29	27	0,2	20e	0,19
26a	0,27	21	0,2	19c	0,29	11	0,19	1c	0,18
18a	0,25	10e	0,2	19a	0,29	6	0,18	26b	0,18
26d	0,25	20a	0,19	24a	0,27	25	0,16	1a	0,17
23	0,25	1c	0,19	28a	0,25	19a	0,15	18a	0,16
19b	0,25	24a	0,19	24d	0,25	2	0,12	5	0,15
18d	0,24	20d	0,19	21	0,24	5	0,12	17b	0,14
22	0,24	4	0,18	26d	0,22	10d	0,12	17d	0,12
10e	0,23	1a	0,18	24c	0,22	19c	0,11	1b	0,1
18c	0,23	28e	0,17	18d	0,21	15	0,1	3	0,1
10b	0,23	3	0,16	17a	0,19	19b	0,08	16d	0,09
19c	0,22	13b	0,16	16a	0,19	7	0,07	18c	0,08
7	0,22	17a	0,15	10e	0,19	26a	0,06	9	0,07
19a	0,22	8	0,15	18c	0,19	22	0,06	24d	0,07
28b	0,22	2	0,15	10b	0,19	16c	0,05	24b	0,05
28c	0,22	25	0,13	23	0,18	21	0,05	18b	0,04
24b	0,21	17c	0,13	17b	0,17	20a	0,03	23	0,04
21	0,21	11	0,13	20e	0,17	28e	0,03	24a	0,04
17d	0,21	16c	0,13	10d	0,17	8	0,03	21	0,03
24c	0,2	18a	0,13	17d	0,15	13b	0	13b	0,03
26c	0,18	19c	0,13	15	0,14	16d	-0,01	26d	0,02
24d	0,18	1b	0,12	26b	0,14	17b	-0,01	20b	0,02
20e	0,17	18d	0,11	18a	0,13	13d	-0,02	10e	0,02
2	0,16	14	0,11	10c	0,13	26d	-0,02	13d	0,02
15	0,15	24b	0,11	28d	0,12	18a	-0,02	26c	0,02
8	0,15	28b	0,11	6	0,1	28b	-0,03	28d	0,01
28a	0,15	17b	0,09	16c	0,09	20b	-0,03	13c	0,01
17c	0,15	6	0,09	17c	0,09	17d	-0,04	28b	0
17b	0,15	28c	0,09	10a	0,09	26c	-0,04	19b	0
1b	0,14	23	0,08	22	0,09	13a	-0,06	24c	0
14	0,13	27	0,08	11	0,08	20d	-0,06	20c	-0,01
16d	0,11	16b	0,06	18b	0,07	18c	-0,06	19a	-0,01
10d	0,1	28a	0,05	7	0,07	24c	-0,08	27	-0,01
16b	0,1	26c	0,05	16b	0,06	17a	-0,08	20a	-0,02

Continúa en la página siguiente

Tabla A.6 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
1c	0,1	20e	0,05	14	0,06	13c	-0,09	15	-0,03
13b	0,09	24c	0,05	2	0,06	16b	-0,09	13a	-0,03
28d	0,09	28d	0,05	3	0,05	20c	-0,1	22	-0,03
17a	0,07	17d	0,04	8	0,04	28a	-0,1	19c	-0,03
3	0,06	9	0,03	5	0,04	17c	-0,1	20d	-0,05
10c	0,06	15	0,03	16d	0,01	24b	-0,1	26a	-0,06
9	0,04	26a	0,03	13d	-0,01	14	-0,1	17a	-0,07
26b	0,04	10c	0,03	13a	-0,01	28c	-0,11	7	-0,09
18b	0,03	18c	0,03	9	-0,03	20e	-0,11	2	-0,09
6	0,02	7	0,03	13c	-0,04	23	-0,11	16b	-0,1
16c	0,01	16d	0,02	4	-0,04	16a	-0,13	10b	-0,1
1a	0	24d	0	1c	-0,05	28d	-0,14	14	-0,1
10a	-0,01	26b	0	1a	-0,11	26b	-0,14	28c	-0,12
11	-0,05	26d	-0,02	13b	-0,18	24d	-0,15	6	-0,13
5	-0,05	10d	-0,03	1b	-0,25	24a	-0,15	28a	-0,15
4	-0,1	10a	-0,04	20b	-0,31	18b	-0,16	8	-0,16
13c	-0,11	5	-0,05	20d	-0,36	18d	-0,17	28e	-0,25
13a	-0,14	18b	-0,06	20a	-0,43	10c	-0,23	11	-0,27
13d	-0,19	22	-0,08	20c	-0,5	10a	-0,24	25	-0,29

Tabla A.7: Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones de Pearson después de una rotación promax

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
28e	0,59	20c	0,85	13a	0,93	18d	0,48	4	0,57
25	0,5	20a	0,73	13d	0,91	16a	0,47	1c	0,55
27	0,47	20b	0,66	13c	0,57	10c	0,39	1a	0,54
19b	0,44	20d	0,57	10b	0,18	24a	0,38	3	0,41
26a	0,44	1b	0,28	19b	0,18	10a	0,37	1b	0,36
19a	0,42	13b	0,22	19a	0,15	24b	0,33	9	0,35
28c	0,42	16a	0,19	21	0,14	26c	0,33	10b	0,32
10b	0,42	13a	0,19	13b	0,14	20e	0,33	10e	0,32
26c	0,42	18a	0,15	17a	0,14	18a	0,31	13d	0,23
24a	0,41	1c	0,15	4	0,14	24d	0,3	13a	0,19
19c	0,4	10b	0,15	11	0,13	17c	0,29	27	0,19
10e	0,38	8	0,14	20d	0,13	23	0,28	19a	0,18
28b	0,37	7	0,14	20b	0,13	18c	0,28	10d	0,17
21	0,36	24a	0,14	16a	0,13	28b	0,27	20a	0,17
24b	0,35	13d	0,14	1a	0,13	28c	0,26	16c	0,16
16a	0,34	2	0,13	24a	0,13	17d	0,26	5	0,14
26d	0,29	28e	0,13	20c	0,13	19b	0,26	19b	0,13
28a	0,29	10e	0,12	8	0,13	17b	0,25	6	0,13
24c	0,28	13c	0,11	10e	0,12	26b	0,25	2	0,13
23	0,28	23	0,11	1c	0,12	26d	0,24	20b	0,11
18a	0,26	1a	0,1	3	0,11	21	0,24	19c	0,11
18c	0,25	14	0,1	2	0,11	26a	0,24	21	0,11
18d	0,25	22	0,1	16c	0,1	24c	0,23	11	0,1
24d	0,25	25	0,1	20a	0,1	27	0,22	15	0,09
17a	0,23	17c	0,1	14	0,1	10d	0,22	18a	0,08
7	0,22	19b	0,09	17c	0,1	16c	0,2	13b	0,08
15	0,22	18d	0,08	28e	0,09	19a	0,2	20c	0,08
2	0,21	21	0,08	1b	0,07	28e	0,2	13c	0,08
17d	0,21	17d	0,08	6	0,07	19c	0,2	25	0,07
17b	0,21	16d	0,08	18a	0,07	28d	0,17	17b	0,06
8	0,2	18c	0,07	25	0,07	28a	0,15	7	0,05
13a	0,2	19a	0,07	24b	0,06	17a	0,14	20d	0,05
22	0,2	3	0,06	19c	0,06	20b	0,14	16d	0,04
20b	0,2	16b	0,06	18d	0,06	10e	0,13	16a	0,04
20e	0,19	26d	0,06	28b	0,05	18b	0,13	26a	0,03
14	0,16	27	0,05	17b	0,05	16d	0,11	22	0,03
3	0,16	15	0,05	16b	0,05	22	0,09	17d	0,02
13d	0,15	9	0,04	28c	0,04	20c	0,09	8	0,02
6	0,15	24c	0,04	28d	0,04	15	0,09	17c	0,02
17c	0,14	20e	0,04	23	0,03	14	0,08	28b	0
10d	0,14	19c	0,04	28a	0,03	13a	0,07	28e	-0,02
11	0,14	17b	0,03	10c	0,01	13d	0,06	18d	-0,02

Continúa en la página siguiente

Tabla A.7 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5
28d	0,13	28b	0,03	20e	0,01	16b	0,05	26d	-0,02
20a	0,13	26a	0,02	24c	0,01	25	0,05	18c	-0,02
16b	0,13	24b	0,02	26c	0	13c	0,05	20e	-0,02
20c	0,13	28d	0,01	9	-0,01	7	0,05	26c	-0,04
1c	0,12	24d	0	27	-0,01	10b	0,04	24b	-0,05
16c	0,07	17a	0	16d	-0,01	5	0,04	23	-0,06
20d	0,07	28a	-0,01	15	-0,01	20a	0,03	24a	-0,06
13c	0,07	16c	-0,01	17d	-0,01	2	0,01	24c	-0,06
16d	0,06	4	-0,01	26b	-0,01	20d	0,01	17a	-0,07
26b	0,06	6	-0,02	7	-0,02	8	0	26b	-0,07
9	0,05	28c	-0,02	18c	-0,02	3	-0,01	14	-0,09
10c	0,04	10c	-0,03	24d	-0,03	13b	-0,01	10c	-0,09
18b	0,01	11	-0,03	26a	-0,04	1c	-0,03	16b	-0,09
1a	0,01	10d	-0,04	10a	-0,05	4	-0,06	10a	-0,1
4	-0,01	18b	-0,04	18b	-0,06	9	-0,08	24d	-0,11
1b	-0,02	26b	-0,06	5	-0,06	6	-0,08	28d	-0,11
13b	-0,02	5	-0,08	26d	-0,07	1b	-0,09	28c	-0,13
5	-0,03	10a	-0,09	10d	-0,08	1a	-0,11	28a	-0,14
10a	-0,07	26c	-0,1	22	-0,13	11	-0,2	18b	-0,15

A.4. Resultados del análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas

Tabla A.8: Valores propios del análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas

factor	Valor propio	factor	Valor propio
f1	7,56	f32	-0,09
f2	4,53	f33	-0,14
f3	3,11	f34	-0,15
f4	2,58	f35	-0,19
f5	2,19	f36	-0,21
f6	2,02	f37	-0,27
f7	1,57	f38	-0,29
f8	1,49	f39	-0,31
f9	1,39	f40	-0,36
f10	1,3	f41	-0,38
f11	1,23	f42	-0,39
f12	1,04	f43	-0,45
f13	0,99	f44	-0,47
f14	0,83	f45	-0,48
f15	0,75	f46	-0,52
f16	0,68	f47	-0,52
f17	0,62	f48	-0,58
f18	0,6	f49	-0,59
f19	0,49	f50	-0,6
f20	0,44	f51	-0,61
f21	0,36	f52	-0,64
f22	0,29	f53	-0,65
f23	0,26	f54	-0,67
f24	0,23	f55	-0,69
f25	0,13	f56	-0,71
f26	0,11	f57	-0,74
f27	0,07	f58	-0,76
f28	0	f59	-0,78
f29	-0,02	f60	-0,79
f30	-0,05	f61	-0,81
f31	-0,06		

Tabla A.9: Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
20c	0,66	26c	0,5	13b	0,92	13d	0,75	1a	0,62	10a	0,46
20a	0,65	28e	0,48	13a	0,72	13a	0,69	4	0,6	10c	0,42
20b	0,65	24d	0,45	13d	0,6	13c	0,62	1c	0,59	18d	0,4
16a	0,52	27	0,44	13c	0,53	19b	0,36	9	0,51	16a	0,4
20d	0,52	28c	0,43	20c	0,43	10b	0,34	3	0,47	16c	0,39
28e	0,46	19c	0,41	20b	0,42	18c	0,33	1b	0,44	20e	0,31
18c	0,45	25	0,39	1b	0,38	11	0,31	10b	0,37	5	0,31
25	0,44	24b	0,37	20d	0,36	17a	0,3	27	0,33	17c	0,28
24a	0,43	26a	0,37	20a	0,35	5	0,28	10e	0,3	4	0,28
27	0,41	28a	0,34	1c	0,29	19a	0,28	19a	0,23	10d	0,27
19b	0,39	26d	0,3	2	0,28	28b	0,28	25	0,23	18a	0,2
18d	0,37	24a	0,3	8	0,25	4	0,27	11	0,22	17b	0,19
10b	0,37	17d	0,29	1a	0,25	10e	0,26	5	0,22	17d	0,17
10e	0,36	21	0,27	21	0,2	28c	0,22	10d	0,21	26b	0,15
26a	0,36	20e	0,26	18a	0,19	16a	0,21	6	0,2	1a	0,15
7	0,36	19b	0,26	19c	0,18	16b	0,21	19b	0,18	18c	0,13
22	0,36	23	0,25	28e	0,18	24c	0,21	19c	0,17	16d	0,12
28b	0,35	17b	0,25	24a	0,16	3	0,2	15	0,16	23	0,12
18a	0,33	19a	0,25	17d	0,13	24a	0,2	26a	0,14	1c	0,1
24c	0,33	10d	0,24	16a	0,12	26c	0,19	16c	0,13	24d	0,1
19a	0,32	26b	0,23	3	0,11	28d	0,18	2	0,13	1b	0,09
26d	0,32	18d	0,23	23	0,11	6	0,18	21	0,1	21	0,09
23	0,32	28b	0,23	17b	0,1	21	0,18	22	0,08	24a	0,08
28c	0,3	22	0,19	10b	0,09	24b	0,17	28e	0,07	18b	0,07
24b	0,29	10c	0,19	17c	0,08	14	0,17	7	0,06	19a	0,06
21	0,28	18a	0,17	24d	0,08	17c	0,17	17b	0,05	3	0,06
19c	0,28	24c	0,17	16c	0,08	18d	0,17	26c	0,05	24b	0,06
17d	0,27	16a	0,16	19b	0,07	28a	0,16	28b	0,03	26d	0,05
26c	0,26	10e	0,16	20e	0,07	25	0,15	13d	0,02	19c	0,04
15	0,26	17a	0,15	19a	0,05	15	0,14	8	0,02	9	0,04
28a	0,24	2	0,15	4	0,05	16c	0,13	26d	0,01	19b	0,04
8	0,23	10a	0,12	25	0,05	28e	0,1	13b	0	28d	0,04
14	0,21	8	0,1	10e	0,05	9	0,09	16d	0	13d	0,03
17c	0,21	10b	0,09	14	0,02	7	0,08	18a	0	28b	0,03
24d	0,2	28d	0,08	27	0,01	26a	0,07	13a	-0,01	20b	0,03
20e	0,2	6	0,06	16d	0,01	1a	0,06	20a	-0,02	26c	0
1b	0,19	15	0,05	18d	0,01	18b	0,05	17d	-0,05	13b	0
16b	0,19	17c	0,05	22	0	19c	0,05	20b	-0,06	13c	0
17b	0,18	16c	0,05	17a	0	18a	0,05	24b	-0,07	13a	-0,01
2	0,18	18b	0,04	10c	0	10c	0,05	17a	-0,07	17a	-0,02
28d	0,17	14	0,03	24b	0	27	0,05	24c	-0,08	20c	-0,02
16d	0,15	18c	0,03	26d	-0,01	17b	0,04	20e	-0,09	22	-0,02

Continúa en la página siguiente

Tabla A.9 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
17a	0,15	13b	0,02	9	-0,04	26b	0,04	28c	-0,1	26a	-0,02
10d	0,13	11	0	26b	-0,05	8	0,04	17c	-0,1	24c	-0,03
1c	0,13	13a	0	10d	-0,05	1c	0,02	23	-0,11	15	-0,04
3	0,1	7	-0,01	11	-0,06	23	0,02	16a	-0,11	27	-0,05
9	0,1	16b	-0,01	6	-0,06	20e	-0,01	18c	-0,11	10e	-0,06
18b	0,07	3	-0,02	26a	-0,06	2	-0,01	24d	-0,12	20d	-0,06
10c	0,06	16d	-0,03	10a	-0,07	16d	-0,02	24a	-0,13	20a	-0,07
6	0,05	13d	-0,05	28a	-0,07	10a	-0,02	28a	-0,13	2	-0,1
26b	0,04	13c	-0,06	28c	-0,09	20d	-0,04	13c	-0,14	16b	-0,1
16c	0,04	1c	-0,09	7	-0,1	20b	-0,05	14	-0,14	14	-0,13
5	0,02	9	-0,2	28d	-0,11	10d	-0,07	28d	-0,15	10b	-0,15
1a	0,02	4	-0,21	16b	-0,11	26d	-0,07	16b	-0,15	28c	-0,15
11	-0,01	1a	-0,21	28b	-0,13	20a	-0,08	20d	-0,15	7	-0,17
13b	-0,01	5	-0,23	26c	-0,13	17d	-0,09	26b	-0,18	28a	-0,17
13a	-0,02	20b	-0,28	15	-0,14	1b	-0,1	20c	-0,18	6	-0,19
10a	-0,02	1b	-0,4	24c	-0,16	24d	-0,12	18d	-0,2	28e	-0,24
13d	-0,06	20c	-0,41	18b	-0,18	20c	-0,12	10a	-0,21	8	-0,28
4	-0,08	20d	-0,45	18c	-0,42	22	-0,25	18b	-0,21	11	-0,39
13c	-0,11	20a	-0,47	5	-0,5	13b	-0,38	10c	-0,26	25	-0,41

Tabla A.10: Columnas ordenadas de la matriz de estructura producida por el análisis factorial estimado con MV sobre la matriz de correlaciones tetracóricas después de una rotación promax

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
25	0,71	20c	0,9	18d	0,64	13a	0,96	4	0,71	5	0,64
28e	0,69	20a	0,88	16a	0,62	13d	0,95	1a	0,71	18c	0,55
27	0,64	20d	0,8	10a	0,49	13c	0,82	1c	0,65	24c	0,36
26c	0,56	20b	0,78	10c	0,46	17a	0,32	1b	0,59	28b	0,33
26a	0,54	1b	0,44	20e	0,43	13b	0,31	9	0,54	16b	0,28
28c	0,52	16a	0,37	16c	0,43	11	0,26	3	0,5	15	0,27
19c	0,52	7	0,34	17c	0,38	19b	0,26	5	0,37	28d	0,27
10b	0,52	13b	0,27	18a	0,38	4	0,24	16c	0,3	7	0,27
28a	0,51	18c	0,26	17d	0,38	21	0,24	10b	0,29	18b	0,26
10e	0,5	8	0,26	10d	0,37	17c	0,23	10d	0,26	11	0,26
19b	0,49	14	0,26	17b	0,37	24a	0,23	10e	0,25	9	0,24
19a	0,48	22	0,25	23	0,36	10b	0,22	19a	0,24	17a	0,24
8	0,46	18a	0,23	24a	0,33	14	0,2	27	0,23	10b	0,23
24d	0,45	16b	0,23	26b	0,33	16c	0,19	19b	0,21	19b	0,23
22	0,44	24a	0,2	24d	0,32	19a	0,19	13d	0,21	4	0,22
24b	0,44	16d	0,2	18c	0,31	8	0,19	15	0,2	18d	0,22
26d	0,41	17c	0,2	5	0,3	3	0,19	19c	0,19	10e	0,21
21	0,38	18d	0,2	24b	0,3	16a	0,18	6	0,16	6	0,19
24a	0,36	23	0,19	21	0,3	16b	0,18	21	0,14	28c	0,19
2	0,36	24c	0,19	16d	0,3	10c	0,18	13a	0,13	16a	0,19
28b	0,36	10b	0,18	18b	0,29	28d	0,18	17b	0,13	14	0,19
11	0,35	15	0,18	26d	0,28	2	0,16	2	0,13	26c	0,18
7	0,34	2	0,17	20b	0,28	6	0,16	20a	0,12	19a	0,18
17d	0,33	28d	0,17	28d	0,27	18d	0,16	18a	0,12	28a	0,17
6	0,32	9	0,15	20c	0,26	24b	0,15	16d	0,12	17c	0,15
23	0,32	10e	0,15	19c	0,25	28c	0,15	16a	0,08	26a	0,13
24c	0,3	1c	0,15	28b	0,24	10e	0,14	11	0,08	24b	0,13
15	0,29	17d	0,15	19b	0,23	20d	0,14	26a	0,08	16d	0,12
17b	0,29	18b	0,15	26c	0,22	1a	0,14	20b	0,07	10a	0,11
10d	0,29	26d	0,13	19a	0,21	28a	0,14	13b	0,07	3	0,1
18a	0,28	28e	0,12	4	0,18	18a	0,13	17c	0,07	10c	0,08
1c	0,27	1a	0,12	24c	0,18	28b	0,12	22	0,07	26b	0,06
3	0,26	28b	0,12	13c	0,17	26b	0,12	28b	0,06	16c	0,06
17a	0,24	13a	0,11	13d	0,16	1c	0,12	20e	0,06	24a	0,05
14	0,24	25	0,11	17a	0,15	17b	0,12	7	0,04	20a	0,04
13b	0,22	19b	0,11	13b	0,15	19c	0,11	10a	0,03	25	0,04
20e	0,21	13c	0,1	26a	0,14	28e	0,11	13c	0,02	20d	0,03
16a	0,2	21	0,1	13a	0,12	20b	0,1	17d	0,02	10d	0,02
16b	0,19	20e	0,09	22	0,09	10a	0,1	25	-0,01	26d	0,02
26b	0,18	17a	0,09	20d	0,07	23	0,09	26d	-0,04	27	0,02
28d	0,18	13d	0,08	20a	0,07	24c	0,08	18d	-0,04	23	0,01

Continúa en la página siguiente

Tabla A.10 – continúa de la página anterior

Ítem	f_1	Ítem	f_2	Ítem	f_3	Ítem	f_4	Ítem	f_5	Ítem	f_6
20b	0,17	19a	0,07	28c	0,07	20e	0,07	20d	-0,04	20e	0,01
18c	0,17	3	0,06	27	0,06	1b	0,07	26c	-0,04	17b	0
9	0,17	17b	0,05	1b	0,05	20c	0,05	18c	-0,05	18a	-0,01
18d	0,17	5	0,05	16b	0,05	26c	0,04	20c	-0,06	13d	-0,01
13a	0,16	28a	0,04	15	0,05	18b	0,03	17a	-0,07	21	-0,03
1a	0,15	26a	0,03	1a	0,03	25	0,02	10c	-0,07	20b	-0,03
16d	0,13	10c	0,02	1c	0,03	16d	0,02	23	-0,08	20c	-0,05
20a	0,08	24b	0,02	14	0,03	20a	0,02	8	-0,08	22	-0,07
13d	0,08	16c	-0,01	3	0,02	17d	0,02	24b	-0,09	13c	-0,08
20c	0,07	6	-0,01	28a	0,01	15	0,01	24c	-0,1	1a	-0,09
17c	0,06	10a	-0,02	10e	0,01	9	0,01	24a	-0,11	28e	-0,09
18b	0,06	27	-0,02	2	-0,02	18c	-0,01	28e	-0,12	17d	-0,1
1b	0,05	28c	-0,03	28e	-0,04	24d	-0,01	28d	-0,12	8	-0,11
16c	0,02	11	-0,03	9	-0,05	5	-0,04	26b	-0,13	1c	-0,12
13c	0,02	24d	-0,07	7	-0,05	7	-0,06	24d	-0,16	19c	-0,13
20d	0	19c	-0,08	10b	-0,12	26a	-0,07	16b	-0,18	1b	-0,16
4	-0,04	26b	-0,09	8	-0,12	27	-0,07	14	-0,18	24d	-0,16
10c	-0,04	4	-0,1	6	-0,14	26d	-0,07	18b	-0,2	13a	-0,16
10a	-0,06	10d	-0,13	25	-0,14	10d	-0,12	28c	-0,2	2	-0,26
5	-0,11	26c	-0,14	11	-0,33	22	-0,21	28a	-0,22	13b	-0,9