PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Pontificia Universidad Católica de Chile

Facultad de Educación

Vacíos de Aprendizaje y Tipologías de Error: Diseño, validación y análisis psicométrico de un instrumento de medición del error mediante el Modelo de Crédito Parcial.

Por:

Camilo Ernesto Lara López

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

Profesora Guía
Verónica Santelices

Noviembre, 2020
Santiago de Chile

©2020 Camilo Ernesto Lara López

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

Dedico esta Tesis a todos los jóvenes de Chile,
deseando que sus equivocaciones sean parte de sus aprendizajes.

Y a mis padres
que siendo muy muy pequeño
les escribí, y les volvería a escribir:

"Queridos mamá y papá
les entrego esto por lo que han hecho por mi;
tengo mis cosas, voy a la escuela
me han cuidado y se han preocupado por mí
y he sentido el cariño de cada uno.

Sé que trabajaron muy duro para esto,
y por eso les entrego esta simple carta
para darles las gracias,
con todo mi corazón."

Agradecimientos

Quisiera agradecer al personal de este Magíster, por brindarme la oportunidad de ingresar a este programa que he disfrutado mucho, y me ha ayudado tanto en mi realización profesional como académica, así como a cada uno de sus Docentes de los cuales aprendí valiosas lecciones.

A la profesora Verónica Santelices por haber considerado ser mi profesora guía; impulsándome siempre más allá de mis propios horizontes, de manera crítica y rigurosa, pero a la vez considerando mis propios ritmos de aprendizaje y descubrimiento. Más que una simple tarea, fue toda una experiencia.

Al Preuniversitario Víctor Jara por su disposición a realizar esta investigación en sus dependencias con el apoyo de su equipo de matemática, en las que mostraron ser un apoyo sincero y desinteresado hacia los estudiantes que buscan cumplir su meta de estudiar una carrera.

Por último, agradezco con especial cariño a mi mamá, por creer siempre en mí y apoyarme incondicionalmente bajo toda cirscunstancia. Especialmente doy gracias a Sofía, por acompañarme en este corto viaje de nuestras vidas, en el que me brindaste tu compañía, comprensión y cariño.

Índice general

Agradecimientos **6**

1. Introducción **11**

2. Objetivos **13**

2.1. Objetivo General 13

2.2. Objetivos Específicos 13

3. Marco Teórico **14**

3.1. Evaluación a gran escala y Edumetría 14

3.2. La educación Pre-universitaria 16

3.3. Uso del error como estrategia pedagógica 17

3.4. Los tipos de error 19

4. Marco Metodológico **22**

4.1. Enfoque Metodológico 22

4.2. Construcción del Instrumento 24

4.2.1. Propósito 24

4.2.2. Restricciones operacionales 24

4.2.3. Constructo o Atributo de interés 24

4.2.4. Subtemas, Distribución y Puntuación de los ítems 25

4.3. Definición de la Muestra y los participantes 27

4.3.1. Población 27

4.3.2. Muestra 27

4.4. Criterios de Calidad 28

4.4.1. Validez de Contenido 28

4.5. El Modelo de Crédito Parcial 28

5. Análisis de Resultados **32**

5.1. Funcionamiento de los ítems 32

5.1.1. Parámetros estimados por ítem 32

5.1.2. Ajuste de los ítems 33
5.1.3. Clasificación de ítems por Thresholds en los CCC 36
5.2. Funcionamiento de los Distractores 39
5.2.1. Comportamiento Esperado 39
5.2.2. Ítems con una Categoría de bajo interés 40
5.2.3. Ítems con dos Categorías de bajo interés 43
5.3. Relación entre Dificultad y Distractores 43

6. Discusión de Resultados **45**
6.1. Dificultad y Tipologías de Errores 45
6.2. Vacíos de Aprendizaje y Tipologías de Error 46
6.3. Tipos de Error en el Contexto Pedagógico 46

7. Conclusiones **48**
7.1. Proyecciones 49

8. Bibliografía **50**

A. Parámetros estimados del ítem **55**
A.1. Dificultad por ítem 55
A.2. Dificultad según tipo de error 56

B. Mapas de Distribución **58**

C. Curvas Características del ítem **62**

D. Curvas de Probabilidad Acumulada **70**

E. Estadísticas Tradicionales del ítem **78**

F. Dificultades de Ítem y Step (Tipos de error) **87**

G. Ítems del Instrumento **89**

Resumen

El presente estudio de investigación tuvo como propósito la construcción, validación y análisis de un instrumento de medición sobre los tipos de errores cometidos por estudiantes de un Preuniversitario chileno en una prueba de alternativas. Se utilizó como referente teórico la Teoría de Respuesta al Ítem, haciendo uso del Modelo de Crédito Parcial, de manera que pudiera obtenerse información tanto del comportamiento de los errores como de la dificultad de los ítems a partir de los resultados.

Para ello, el estudio se enmarcó dentro de un enfoque psicométrico de investigación cuantitativa. Contando con la participación de 105 estudiantes que respondieron el test, los cuales estaban en su mayoría entre los 17 y los 19 años de edad. Los datos se obtuvieron a través de la plataforma Google Form en el que los estudiantes respondían de forma online en la misma sala de clases, y luego analizados a través del software CONQUEST 2.0 (2007).

Los resultados obtenidos dan cuenta, de una posible relación entre la dificultad del ítem y la probabilidad de cometer determinados tipos de error. Esto podría favorecer una comunicación horizontal entre la psicometría y la edumetría al utilizar herramientas estadísticas para obtener resultados que se pueden interpretar mediante distintas tipologías de errores en diversos niveles o contenidos matemáticos. Las conclusiones del estudio aportan sugerencias y recomendaciones para poder aplicar este análisis en evaluaciones de aula o a distancia, con énfasis en la retroalimentación de los estudiantes.

Abstract

The purpose of this research study was the structure, authentication and analysis of a measuring instrument about kinds of mistakes made by students of a Chilean Pre-University level with a test involving alternatives. The Item Response Theory (IRT) was used as a theoretical reference, making use of the Partial Credit Model, so that information could be obtained from both the behavior of the mistakes and then the difficulty of the items from the results.

For that purpose, the study was framed within a psychometric approach of quantitative research. It involved 105 students who answered the test, most of them were mostly between 17 and 19 years of age. The data were sought through the Google Form platform in which students answered, in the same way as in online mode, but in classroom and then they were analyzed through the software CONQUEST 2.0 (2007).

The results features results obtained in a possible relationship between the difficulty of the item and the probability of making certain mistakes. This could encourage a horizontal communication between psychometrics and edumetrics by using statistical tools for obtaining results that can be interpreted by different types of mistakes at various levels or mathematical content. The conclusions of the study contribute suggestions and recommendations to be able to apply this analysis on classroom or remotely assessments, that emphasized feedback student interaction.

Capítulo 1

Introducción

En este trabajo se utiliza una forma de incluir algunos de los principios de las buenas prácticas de retroalimentación propuestas por Nicol y Macfarlane (2006), a través de la creación de un instrumento de medición que analice los errores frecuentes cometidos por un grupo de estudiantes en el desarrollo de habilidades matemáticas presentes en el álgebra lineal, para posteriormente ser analizado mediante el Modelo de Crédito Parcial evaluando su utilidad y practicidad para el mejoramiento del aprendizaje de los estudiantes, y evaluar si este Modelo es aplicable a la realidad que enfrentan.

Actualmente existen nuevos Modelos de análisis de evaluaciones psicométricas provenientes del Modelo de Rash, como el Partial Credit Model, que permiten obtener información cualitativa de estos test e incluso mapear el nivel de habilidad de los estudiantes en relación al nivel de dificultad de los ítems y sus alternativas, a modo de obtener información fiable que permita una retroalimentación o ajuste personalizado de la enseñanza. Una de las ventajas de este Modelo, es que permitiría obtener información acerca de cuál es el tipo de error que está cometiendo el estudiante al contestar un ensayo PSU (por ejemplo), para que pueda trabajarlo y mejorarlo, lo que vuelve necesario evaluar si este modelo es aplicable a la realidad que enfrentan los estudiantes, donde muchas veces deben rendir ensayos que se vuelven un puntaje indescifrable (en ocasiones transformado a puntaje PSU) donde no se considera un relato coherente entre lo logrado y las dimensiones críticas de los objetivos por lograr (Hattie y Timperley, 2007).

Parte de las limitaciones del estudio, fueron la alta demanda laboral docente y el gran número de evaluaciones exigidas, lo que hizo que se mantuviera una tendencia hacia la automatización de la toma de datos a modo de interrumpir lo menos posible los procesos pedagógicos. También, existieron ciertos eventos sociales en el País que permitieron acceder sólo a una muestra limitada del total de sujetos. Por otro lado, la institución correspondía a un preuniversitario popular, en el que asistían sujetos

en los que por declaración de la coordinadora, muchas veces no terminaban la secundaria en buenas condiciones de rendir la PSU, siendo una característica que no se consideró para este estudio, pero que tal vez pueda considerarse en futuros estudios por ejemplo, incluyendo a estudiantes que asistan a otros preuniversitarios, a través de algunas escuelas. Finalmente, aún existe un arduo debate en torno al objetivo y usos de la evaluación con estos nuevos modelos, en los cuales se está intentando "encontrar un balance entre las evaluaciones sumativas y formativas a través del fortalecimiento de los procesos de evaluación" (Organización para la Cooperación y Desarrollo Económico OECD, 2013), lo que dificultó algunas decisiones en la construcción del instrumento, como si es mejor enfocarse en abarcar los contenidos o las competencias esenciales para el buen desempeño de los estudiantes en la educación superior (como declara el Ministerio de Educación para la nueva Prueba de Transición Universitaria), o si el proceso de preparación para esta prueba debe incluir una retroalimentación mas orientada hacia los puntajes o hacia los resultados cualitativos evidenciados.

El presente estudio está organizado comenzando por mostrar sus objetivos, para luego pasar al Marco Teórico, donde se abordan los temas de evaluación a gran escala y a pequeña escala, también los orígenes de la educación Preuniversitaria y parte de su metodología utilizada, para luego terminar con distintos aspectos acerca de la inclusión del uso del error en la enseñanza. Posteriormente, el estudio sigue con el marco metodológico en el que se profundiza sobre algunos aspectos de la construcción del test como la asignación de puntaje para cada uno de los tipos de errores, la definición de la muestra y el modelo utilizado para el análisis de los datos. Luego se observa el análisis de resultados, junto con algunas figuras que permiten una visualización mejor de los datos. Para terminar con la discusión de los resultados respecto a tres puntos fundamentales y las conclusiones del proyecto.

Capítulo 2

Objetivos

2.1. Objetivo General

Construir, validar y analizar mediante Modelo de Credito Parcial un instrumento de detección de errores conceptuales en el álgebra lineal, en estudiantes que están en preparación para rendir la PSU a través de un Preuniversitario.

2.2. Objetivos Específicos

- Sistematizar los avances teóricos en la detección de errores conceptuales en matemática, como herramienta evaluación diagnóstica y formativa.
- Construir una medición sobre los tipos de errores cometidos, en la aplicación de las propiedades del álgebra lineal.
- Caracterizar la relación entre los índices de habilidad y la probabilidad de cometer un determinado tipo de error, a través de las Curvas Características del Ítem.

Capítulo 3

Marco Teórico

Este capítulo tiene como propósito describir en cuatro apartados el marco teórico utilizado para esta investigación. En el primer apartado se describen algunos aspectos sobre las evaluaciones a gran escala en Chile, y sus diferencias con respecto a la evaluación a nivel de sala de clases, la retroalimentación y el uso del error.

En el segundo apartado, se describe la relación entre Prueba de Selección Universitaria y los Preuniversitarios, así como la forma en que estos han ido adaptando sus contenidos o metodologías de acuerdo a los cambios de esta prueba.

En el tercer apartado se profundiza en el uso del error como estrategia pedagógica a través de la evaluación, también en la manera de medir distintos tipos de error a través de una prueba de alternativa.

En el cuarto apartado, se muestran los diferentes tipos de error caracterizados desde distintos puntos de vista (características de desarrollo, de orígen epistemológico, origen en obstáculos, ausencia de sentido, lenguaje algebraico, etc.) y definidos según su relación con el contenido escogido.

3.1. Evaluación a gran escala y Edumetría

En los principios de la evaluación a gran escala en Chile, los usos previstos de una evaluación llamada "validez consecuencial" no fueron considerados debido a que fueron concebidas en una época de industrialización en las últimas décadas del Siglo XX, bajo un paradigma de mercado o de modelos de elección (UNESCO, 2004) que fue permeabilizando algunos aspectos y mecanismos de gestión de la producción en masa, a la "producción" y gestión de la formación de personas educadas, tales como los horarios preestablecidos, la separación por asignaturas, el ingreso en base al año de nacimiento, la evaluación como un índice de calidad netamente cuantitativo, entre otros.

En este contexto, surge la la Prueba de Selección Universitaria en el año 1967

con la ayuda de Leyton Soto y Erika Himmel König (Escudero, 2015) como una solución al sistema de admisión Universitaria mediante Bachillerato y evaluaciones específicas por carrera y Universidad (DEMRE, 2017). Una prueba estandarizada que permitió eliminar la subjetividad con la que se evaluaba el nivel académico en el Bachillerato o en otras instancias de evaluación, alejándose de la "Evaluación para el Aprendizaje", como lo muestra el estudio de "Evidencia Preliminar para la reducción de contenidos en la PSU de Matemáticas", que da cuenta de las dificultades y cuestionamientos actuales sobre esta prueba, sobre aspectos de validez de contenido[1], instruccional[2] y predictiva[3] (Farías y Carrasco, 2012), debido en mayor medida a las diferencias entre los objetivos y metodologías de los establecimientos Municipales, Subvencionados y Particulares. Es así como también el SIMCE y otras mediciones estandarizadas, se han vuelto el objetivo de aprendizaje de muchos establecimientos educativos modelando sus prácticas educativas y desaprovechando el potencial de este tipo de mediciones, que pueden servir de insumo para mejorar los aprendizajes a través de evaluaciones formativas y la detección temprana de errores, en la educación formal.

Dicho de otra forma; el objetivo de la enseñanza se ha vuelto la medición, en vez de que el objetivo de la medición, se haya vuelto la enseñanza.

El uso de la medición para propósitos formativos relativos al "beneficio" que le brinda al estudiante para su aprendizaje, es un tema que aún no encuentra consenso dentro de la comunidad científica , ya que aún existen diversos aspectos, como por ejemplo, los distintos puntos de vista de la validez en evaluaciones a gran escala y a nivel aula, que requerirían un cierto tipo de integración para poder aplicar una medición estandarizada a una sala de clases, considerando además, que la mayoría de los docentes no posee la formación en evaluación suficiente ni el tiempo para ese trabajo (Forster, C.; 2017). Es por ello, que paulatinamente se ha buscado incorporar los aspectos edumétricos mencionados por Brookhart (2003), de manera que el foco deje de estar tanto en la medición y se valla incorporando más hacia la enseñanza, mediante nuevas directrices como las que están adoptando la Agencia de Calidad de la Educación (2018), el Tercer Estudio Regional Comparativo y Explicativo (TERCE) (Treviño et al., 2016), el TIMSS, entre otros.

[1]Como los establecimientos TP no están orientados a la educación superior, pueden tener planes de estudio más restringidos incluyendo sólo algunos de los objetivos fijados por la PSU (Farías y Carrasco, 2012).

[2]Las diferencias instruccionales entre los establecimientos TP y los subvencionados o particulares, podrían deberse a que dedican más tiempo al desarrollo de competencias para la especialidad que al desarrollo de habilidades académicas (Farías y Carrasco, 2012).

[3]En otros estudios (como el Informe Preliminar para la Reducción de Contenidos de la PSU Matemática del DEMRE) se ha demostrado que la validez predictiva de la PSU para el rendimiento académico del 1er año de estudios superiores, sólo es significativo para la PSU de Matemáticas con una magnitud inferior a 0,38, siendo el NEM un mejor predictor en este ámbito.

Cabe destacar que una prueba con alternativas puede fácilmente entregar información relacionada al nivel de la tarea; como cuántas respuestas están correctas e incorrectas, la posición del estudiante con referencia a un grupo objetivo o algunos indicadores sobre el desempeño de los reactivos de la prueba. Esto es considerado por Hattie y Timperley (2007) como una retroalimentación centrada en la tarea, que si bien, permite conocer el tipo de situaciones en las que un estudiante se desenvuelve exitosamente, no brinda información sobre su desempeño en los distintos procesos involucrados en las situaciones calificadas como "incorrectas", a diferencia de la retroalimentación efectiva que es un proceso que necesita proveer información relacionada específicamente a la tarea o al proceso de aprendizaje, que "rellene la brecha entre lo que se sabe y lo que falta aprender" (Sadler, 1989).

Esto fue considerado por Brookhart (2003) como una de las ocho principales deficiencias que le atribuye a las mediciones psicométricas al considerarlas como parte de una evaluación a nivel de aula, acerca del "beneficio" que recibe el estudiante de una evaluación, ya que pueden proveer información válida y confiable en términos psicométricos, pero no le brindan un insumo que le permita mejorar su aprendizaje, teniendo de hecho, criterios de validez para evaluaciones a gran escala y a nivel aula diferencias casi opuestas en su mayoría (Forster, C.; 2017).

3.2. La educación Pre-universitaria

Una clasificación amplia de los tipos de educación corresponde a la Educación Formal Planificada, la No Formal Planificada y la Informal No Planificada. Las dos últimas se caracterizan por no estar sujetas a las leyes de la República, un ejemplo de estas son los Preuniversitarios y las clases particulares respectivamente (Revista de Enfoques Educacionales 13 (1), 2016: 91-102). Los Preuniversitario se iniciaron al mismo tiempo que la Prueba de Aptitud Académica (P.A.A.) en 1967, pero a través de la contratación de profesores particulares, que posteriormente fueron realizando una estructuración propia de los contenidos evaluados en la P.A.A.

En el 2003, este test de admisión pasó a llamarse Prueba de Selección Universitaria o PSU, cuya principal diferencia estaba en que su objetivo ya no era "predecir el rendimiento académico en la universidad, (sino) medir la calidad de la enseñanza media" (Morgado, 2016). Esto forzó a los Preuniversitarios a adaptar los contenidos curriculares del Ministerio de Educación, a una metodología propia, pero sin las presiones sociales y administrativas de la enseñanza formal, lo que brinda una oportunidad para la investigación de estrategias evaluativas no calificadas, pese a que el enfoque esta en la preparación para contestar las preguntas de las pruebas y no la formación de personas (Escudero, 2015).

La estructura rígida de las metodologías de algunos preuniversitarios, produce un efecto de "selección natural" en los estudiantes que pueden seguir un alto rendimiento, ya que la enseñanza no se adapta a los enseñados. Esto sugiere enfocarse en los Preuniversitarios que adaptan su ritmo de enseñanza al nivel y avance de los estudiantes, enfocándose en sus propios objetivos, creando el ambiente ideal para la detección e implementación de estrategias de levantamiento de información para la toma de decisiones sobre la enseñanza, y entrega de retroalimentaciones asertivas.

3.3. Uso del error como estrategia pedagógica

"El Error es un elemento inherente a cualquier proceso de aprendizaje"
(González, 2015)

Chile se destaca en los aspectos relacionados a la organización del aula, desde el manejo de la conducta hasta la planificación de las clases, llegando a obtener un 6,5 (en escala de 1 a 7) en estos aspectos (Taut, Manzi, y Molina, 2013). Sin embargo, se dejan muy de lado los relacionados al "Apoyo Pedagógico", donde se encuentra el desarrollo de conceptos y la calidad de la retroalimentación. Esto se confirma al observar el porcentaje de profesores que aplican Buenas Prácticas Pedagógicas en estrategias de enseñanza y ambiente de clases (Preiss et al., 2014), donde la explicitación de los procesos de razonamiento es utilizado por un 64,2 % de los profesores, relegando la evaluación formativa y uso del error a un 11,3 % y 7,3 % respectivamente:

Porcentajes de Profesores que Implementan (BPD) en Estrategia y Ambiente (N = 151)

Dimensión y práctica	%
Estrategia	
Explicitación de los procesos de razonamiento	64.2
Relacionar con lo cotidiano	42.4
Uso creativo de recursos	33.8
Relacionar con contenidos previos	25.8
Trabajo colaborativo	16.6
Evaluación formativa/procesual	11.3
Uso del error	7.3
Estrategias de resolución de problemas	4.6
Integración de contenidos	3.3

Figura 3.1: Porcentaje de implementación de estrategias.

Se puede observar en esta tabla que las estrategias más utilizadas son las que necesitan una menor inversión de tiempo por parte del profesor, o las asociadas a

la planificación. De hecho existe en los docentes de matemática una tendencia a identificar los errores habituales en la resolución de problemas, reconocer sus causas y proponer una manera de superarlo en los estudiantes (Santagata, 2005; Son, 2013; Son y Sinclair, 2010). Sin embargo, cada estudiante puede estar en un tipo de error diferente y el profesor suele no tener suficiente tiempo para reconocer los diferentes tipos de errores de cada estudiante (Hattie y Timperley, 2007), más aún considerando que cada instancia de "evaluación" debe ser aprovechada para cumplir administrativamente con la alta cantidad de notas exigidas por los establecimientos. Sin embargo, no se ha encontrado evidencia de alguna ley, mandato o similar en el que se exija a los establecimientos educativos el uso de esta regla. Más aún, en la nueva reforma evaluativa para el año 2020, se menciona expresamente que el número de calificaciones dependerá de las necesidades de la planificación de la enseñanza[4].

Es así como teniendo la oportunidad de tomar decisiones pedagógicas para la enseñanza, lo que se necesita es de instrumentos que permitan el levantamiento de información de forma masiva, pero enfocados en el nivel de aprendizaje, más que en determinar cuántos ítems contestó correctamente un estudiante, como lo sería un test dicotómico. Las pruebas de alternativas (más utilizadas para la preparación de la PSU), pueden cambiar el enfoque con el que el estudiante aprende, hacia aspectos más técnicos y superficiales que hacia un aprendizaje profundo del conocimiento (Bloxham & Boyd, 2007), esto incurre en la necesidad de evaluaciones que centren su mirada en la detección de niveles de errores de manera específica, para que puedan ser superados. Para ello es necesario anticiparse a la diversidad de errores que pueden ser cometidos e integrarlos a la evaluación de forma politómica, y a la enseñanza basándose en dichas respuestas (Gonzalez, 2015).

Sin embargo, gran parte de la literatura encontrada sobre el uso del error, lo aborda desde la metodología de enseñanza más que desde la evaluación (González, Gómez & Restrepo, 2015; Rosso & Barros, 2013; Sánchez & López, 2009; Villaroel, et al., 2015), lo que coincide con la escasa literatura sobre el uso que le dan los profesores al error en sus prácticas docentes (González et al., 2015), a pesar de que el efecto que este tiene es positivo y significativo en diversas investigaciones (Keith y Frese, 2008).

Esto abre un campo de investigación cuya relevancia radica en el uso educativo que se le puede brindar a la evaluación del error, como aporte a las estrategias de enseñanza o paradigmas educativos que consideran evaluaciones formativas como parte de su proceso, considerada una metodología de alto impacto[5], con bajo costo

[4]Anteriormente era usual colocar tantas notas como horas tenía la asignatura a la semana, además de otras exigidas por el establecimiento.

[5]De acuerdo al sumario del tamaño del efecto relacionados al feedback, en un metanálisis realizado por Hattie y Timperley (2007).

de implementación y abundante evidencia que la respalda[6].

3.4. Los tipos de error

Desde la epistemología del conocimiento de Bachellard y la Teoría de las Situaciones Didácticas de Brousseau (2001), Socas (1997) y Radatz (1979), esquematizadas por Rosso (2013) para la confección de una "Taxonomía de errores en el aprendizaje de espacios vectoriales", es que se ha ido cambiando a lo largo del último tiempo la forma en que entendemos el error, ya no como parte de la ignorancia de una persona, sino como parte de un conocimiento previo que ha tenido buenos resultados en un contexto diferente al que está siendo aplicado ahora (D´amore, 2014, p. 63). Es así como bajo esta nueva perspectiva, que el error constituye un conocimiento válido para el estudiante, en el que en vez de estar equivocado este ha realizado un "uso inadecuado del conocimiento", aludiendo a las múltiples formas de ser utilizado más que a si está correcto o parcialmente incorrecto, de donde pueden ser medidos diferentes tipos y niveles de errores a través de una evaluación psicométrica. Es por ello que se entenderá el error u "obstáculo cognitivo", como la transición de un conocimiento (factual, conceptual o procedimental) implicado en dar una respuesta frente a una situación de desempeño, desde un contexto en el que es válido a uno en el que resulta inadecuado.

Este concepto de error fue concebido por Bachellard hacia el 1938, como el último obstáculo epistemológico del conocimiento, llamado obstáculo del conocimiento cuantitativo, que son barreras intelectuales que se oponen a la formación de un espíritu científico y que "dificultan el conocimiento certero de lo real" (Villamil, 2008) y fue trasladado al campo de la educación por Brousseau en la década del 90' (Tall, 2014). Desde esta perspectiva, Tall (1989, citado en Castro, 2000, p.136) lo menciona como "obstáculos cognitivos" distinguiendo dos categorías; las basadas en la secuencia de un tema como la graduación de la complejidad, y las basadas sobre casos simples, como limitaciones de la enseñanza en casos simples por largos períodos antes de internarse en casos más complejos. A pesar de ello, Socas coloca estas dos categorías de Tall dentro de los obstáculos de orígen didáctico, que son resultado de las elecciones didácticas que se realizan al establecer un proyecto educativo (Rico, 2000). A esto agrega dos tipologáas de obstáculos que define como de origen ontogénico o psicogénico (debido a las características del desarrollo del niño) y de origen epistemológico (relacionados con el propio concepto). Es así como podríamos clasificar en base a estas ideas, algunas categorías para los errores:

[6]En base a la catalogación realizada por SUMMA (Laboratorio de Investigación e Innovación en Educación para América Latina y el Caribe), disponible en https://www.summaedu.org/retroalimentacion-formativa/

- Características del Desarrollo del Niño
  - Ontogénico
  - Psicogénico
- De orígen Epistemológico (Obstáculos Cognitivos)
  - Graduación de la Complejidad
  - Permanencia de casos simples

Con relación a los de orígen epistemológico, la permanencia de casos simples tiene mayor relación con el tipo de deciciones metodológicas que toma el profesor acerca de su enseñanza, siendo un tema para abordar más bien desde la didáctica, pero la Graduación de la Complejidad es un tema que puede abordarse desde la psicometría de acuerdo a la forma en que se mide y valora cada uno de los conocimientos. Bajo esta perspectiva, el error constituye un conocimiento válido para el estudiante, que en vez de estar equivocado, ha realizado un "uso inadecuado del conocimiento", el cual puede tener diferentes grados de inadecuación, lo que es lo mismo, distintos tipos de error frente a un mismo problema. Siendo así, una manera de adquirir evidencia de esto es a través de una evaluación politómica (ya que son varias respuestas para una misma pregunta), con motivo de dar cabida y valorar los diferentes "grados" de error que podría tener un estudiante, pudiendo considerar algunos de los siguientes tipos de errores descritos por Socas:

1. Orígen en Obstáculos
   a) Clausura
   b) Proceso o Producto
   c) Yuxtaposición
   d) Concatenación
2. Ausencia de Sentido
   a) Orígen en la Aritmética
   b) Procedimientos
3. Lenguaje Algebraico
   a) Tautologías
   b) Veracidad

De estos, se consideraron dos que permitirían mostrar evidencias acerca de errores sistemáticos asociados a la confusión (Yuxtaposición) e invención de propiedades relacionadas con el álgebra lineal, descritas a continuación:

Yuxtaposición : **Error Moderado**. Combinación o recombinación de propiedades algebraicas utilizadas en situaciones de características similares, para las que la propiedad algebraica original estaba definida.

Concatenación : **Error Complejo**. Extensión de propiedades a situaciones de mayor complejidad o extensión, pudiendo complejizar o inventar una propiedad para adaptarla a una situación más compleja.

Estos tipos de error, son factibles de cometer frente a una misma situación problemática, lo que permite la construcción de un instrumento de medición politómico que considere estos errores dentro de sus alternativas, con diferentes graduaciones en sus puntajes, para ser analizado posteriormente mediante el Modelo de Crédito Parcial, de manera que deje de lado las connotaciones negativas sobre el error cuando la respuesta no es la esperada.

Capítulo 4

Marco Metodológico

Este capítulo tiene como propósito describir en tres apartados la metodología utilizada para esta investigación. En el primer apartado se describe el enfoque metodológico y el paradigma en el que está inmerso el estudio, así como los aspectos de los resultados en los que se profundizó y algunas características propias del estudio que lo identifican dentro de ese paradigma.

En el segundo apartado se describe como fué realizada la construcción del instrumento, considerando las 7 fases de construcción de un test tradicional mencionadas por Arías (1996), que corresponden a definir el propósito con el que se usarán las puntuaciones del test, luego identificar las restricciones con las que deberá operar, el atributo de interés, las especificaciones de contenidos, el formato de los ítems, un plan para la selección de los ítems más adecuados y la especificación de los valores estadísticos que se calcularán.

En el tercer apartado, se explica cómo se llevó a cabo el análisis tanto del instrumento completo, como de sus ítems de manera individual, considerando cada una de sus alternativas de respuesta a través del Modelo de Crédito Parcial. Para ello se trabajó con distintos estadísticos obtenidos por el software CONQUEST 2.0 como la bondad del ajuste del ítem a través de los Weighted, Unweighted Mean Square y Discriminación, y la dificultad de los mismos a través de los valores de sus Thresholds, que se asociaron a las curvas de probabilidad acumulada de los ítems. Finalmente, termina con el análisis del funcionamiento de los distractores a través de los valores delta (δ) en las curvas características de los ítems y las curvas de probabilidad acumulada.

4.1. Enfoque Metodológico

El presente estudio, se enmarca en el paradigma cuantitativo de investigación, dado que pretende realizar una medición válida y fiable de un constructo a través de

información objetiva y cuantificable, profundizando en un aspecto específico de los participantes bajo ciertas normas estadísticas (Cook and Reichardt, 1986). Además, se considera un tipo de investigación no experimental dado que no involucra una manipulación de las variables (Ritchey, 2004) y no busca influir sobre los participantes, sino conocer el estado o nivel que poseen sobre un constructo.

Los datos fueron obtenidos a través del uso de un test que permitió evaluar el nivel de habilidad de los participantes frente a diferentes situaciones problemáticas, relacionando entre sí sus propios niveles de habilidad como los distintos niveles de dificultad de cada situación. Es así como el uso de este enfoque, permitió la evaluación de distintos modelos de medición estandarizados basados en la psicometría, de modo que pudiera ser elegido el más adecuado para los fines de este estudio (Masters & Right, 1982). De esta manera, el estudio se planteó como el análisis de los resultados de un test basándose en el Modelo de Crédito Parcial, e incluyendo juicio de experto y entrevistas cognitivas, con graduación del puntaje en sus alternativas, debido a que permitía analizar de manera más específica el comportamiento de cada una de las alternativas de los ítems (Wilson, 2005).

El grupo en el que se aplicó, son estudiantes de un Preuniversario popular de la Región Metropolitana el cual tiene el propósito de ayudar especialmente a los jóvenes que han tenido malas bases de enseñanza secundaria, y que, por lo tanto, les interesa la búsqueda de métodos que permitan determinar los conceptos específicos en los que estos estudiantes poseen deficiencias, lo que ratifica la pertinencia del uso de este modelo para este contexto en particular, sin quitar que posteriormente pueda ser utilizado para otros contextos con fines de retroalimentación pedagógica. Las entrevistas cognitivas a los jueces expertos (particularmente a los profesores del preuniversitario) fueron de particular importancia para definir estos límites entre los posibles resultados de una evaluación a gran escala o a nivel aula, que eran de interés para los profesores del preuniversitario.

El instrumento fue diseñado en base a algunos requerimientos de tiempo, contenidos propios del preuniversitario y formato de adquisición de respuestas necesarios para la ocasión; la disponibilidad de tiempo estaba estrechamente relacionada con el permiso de ocupación de la sala por lo que se ajustó a los parámetros del preuniversitario, los contenidos escogidos tuvieron una confirmación por parte de los profesores del preuniversitario respecto a la pertinencia de su inclusión en el test y su opinión sobre el nivel de los estudiantes frente a los mismos, finalmente la forma de adquirir las respuestas se decidió realizarla mediante medios digitales (Un link en una página web o mediante código QR) dada la costumbre de los estudiantes a estos medios y a la facilidad con que permite su posterior análisis.

4.2. Construcción del Instrumento

4.2.1. Propósito

En una entrevista con los docentes del Preuniversitario, ellos describieron que la preparación para la PSU puede llegar a ser compleja debido al tiempo que poseen para la enseñanza de cada uno de los contenidos, y por lo mismo, una retroalimentación individualizada o una adaptación de la clase según el dominio que posee cada estudiante sobre algún concepto en particular es aún más difícil, es por ello que buscan una forma de ir mejorando sus diagnósticos, como en un principio podrían ser los Test Referidos al Criterio o "TRC" (Martinez, 1996). Por lo mismo, uno de los propósitos del test, es la obtención de información que permita identificar deficiencias o errores particulares de rendimiento de manera rápida y útil para los docentes, como una forma de apoyar la adaptación de sus clases a la realidad de los estudiantes.

4.2.2. Restricciones operacionales

Por solicitud de los profesores, el test no debía durar más de 30 min para no interrumpir el normal desarrollo de las clases, por lo cual se realizó un test que poseía preguntas en las que se esperaba que pudieran ser respondidas en menos de 1 minuto ya que correspondían sólo a la habilidad de identificar, dejando una minoría de preguntas de mayor complejidad.

El lugar donde se realizaría, sería en la misma sala donde hacen clases normalmente, durante las clases de matemática. Además, como inicialmente se pretendía realizar el test a todas las sedes del preuniversitario, se modificó de modo que pudiera ser respondido online a través de una encuesta realizada en Google Forms, así el test podría compartirse con los docentes y con los estudiantes a través de un link a una página web que contenía el vínculo, y además, a través de un código QR legible a través de la cámara de cualquier teléfono celular, que direccionaba al mismo test. Para los estudiantes que no poseían señal de internet se llevaron algunas copias impresas del test.

4.2.3. Constructo o Atributo de interés

Sobre el contenido específico a abordar en el test, se les consultó sobre los conceptos en que ellos presenciaban mayor dificultad en la enseñanza a modo de un análisis por Juicio de Expertos, llegando a la conclusión de que el álgebra era uno de los más difíciles e importantes debido a que de estos conceptos depende la posterior comprensión y aplicación en el campo de la geometría y sobre todo en la Estadística. También se realizó una revisión de otros instrumentos de similares caracerísticas, las mismos

test que realizó el Preuniversitario y algunas investigaciones publicadas dentro de las cuales destacó la definición propuesta por la NCTM (2011) debido a que aborda la construcción de los objetos matemáticos desde una perspectiva pedagógica, con lo cual se definió el constructo como la *Comprensión interconectada de expresiones algebraicas, ecuaciones de primer grado y gráfico de funciones del álgebra lineal de la enseñanza secundaria.*

4.2.4. Subtemas, Distribución y Puntuación de los ítems

De manera más específica, se desglosaron los contenidos a partir de las definiciones que la NCTM (2011) brinda a los subtemas abordados, y que menciona como "Big Ideas", las que fueron contrastadas con el Currículum Nacional:

Expresiones : Unidad mínima que permite trabajar con ecuaciones y funciones, se caracteriza por poseer términos algebraicos compuestos por coeficientes y valores desconocidos.
Ítems: 6, 15, 22 y 25.

Variables : Es una herramienta para expresar ideas matemáticas de forma clara y concisa. Es un valor desconocido al que se le asigna una letra para representarlo, o valores que cambian bajo un patrón determinado o que se intenta conocer.
Ítems: 3, 4, 14, 20 y 23.

Igualdades : Expresión matemática que indica la equivalencia entre 2 expresiones algebraicas que utilizan una misma variable.
Ítems: 1, 2, 8, 9, 10, 11, 12, 13, 24, 26 y 27.

Funciones : Son asociaciones entre variables que describen situaciones de cambio o de relación numérica, y que pueden representarse de forma simbólica, gráfica o numérica.
Ítems: 16, 17, 18, 19 y 29.

Resolución de Problemas : Corresponde a la resolución de ecuaciones a través de algoritmos generales o alternativos, haciendo uso de propiedades algebraicas.
Ítems: 5, 7, 21, 28 y 30.

Como formato de respuesta para cada uno de los ítems se escogió el de selección múltiple, dado que los otros formatos que propone Bennet (1993) tales como ordenación de un material, sustitución, completación, construcción o presentación (que podrían fomentar la medición de otras habilidades como Modelar, Representar, Resolver Problemas y en especial Argumentar y Comunicar) no son compatibles con

la tecnología de Google Forms, que es con la que se construyó el instrumento.

Además, se consideró como parte del estudio los errores asociados a razones no espistemológicas, es decir, no relacionados a obstáculos cognitivos, si no más bien, a errores de cálculo en los que las propiedades para la resolución del problema están bien aplicadas, pero el estudiante (debido a razones de apuro, stress, u otras) comete algunas equivocaciones en su aplicación u operaciones aritméticas involucradas. Esto permitió que las puntuaciones fueran asignadas de manera gradual de modo que a la alternativa correcta se le asignaban 3 puntos y a la alternativa más incorrecta se le asignaban 0 puntos, asignándoles puntajes de 1 y 2 puntos a las alternativas con categorías de error intermedias. Dichas puntuaciones se basaron en los tipos de error más frecuentes valoradas de acuerdo a la Tipología de errores de Socas (1997), como se indica en el Cuadro 4.1.

Cuadro 4.1: Puntuación de las alternativas

Alt	Tipología	Pje	Justificación
a)	Concatenación	0	**Error Complejo:** Forma compleja de error, donde se inventa una propiedad o método con base en anteriores.
b)	Yuxtaposición	1	**Error de Superposición:** Forma de mediana complejidad, en la que el estudiante utiliza una propiedad conocida o familiar, en una situación donde no aplica.
c)	No Epistemológico	2	**Error de Cálculo:** Forma más "simple" de error, debido a algún hecho fortuito de parte del estudiante como un error de cálculo o cambio menor en las características de las propiedades.
d)	Respuesta Correcta	3	**Error Ausente**: El estudiante aplica las propiedades y cálculos necesarios de manera correcta para la resolución del problema presentado.

Para cada uno de los ítems, se tuvo que agregar una quinta alternativa de respuesta debido al formato utilizado por el Preuniversario, lo que permitió a su vez dar una opción adicional con algún tipo de error en particular. El foco en el que se puso mayor importancia en la construcción de los ítems, fue en la justificación de sus alternativas, de modo que no fueran sólo distractores visuales, sino posibles caminos alternos por parte de los estudiantes, como se muestra en la Figura 4.1.

Figura 4.1: Ejemplo de ejercicio y sus justificaciones

Subdimensión	Nº de Construcción	Reactivo	Alternativa		Puntaje	Resp. Correcta	No epistemológico	Yuxtaposición	Concatenación	Posible Razonamiento	Justificación
S_1	1	¿Cuál de las siguientes alternativas es una forma equivalente a la expresión $\frac{p+q}{p-q} : \frac{p-q}{p+q}$?	a)	0	2		X			$\frac{p+q}{p-q} - \frac{p+q}{p-q} = 0$	Invierte el divisor como en una división de fracciones, pero resta las fracciones en vez de multiplicarlas.
			b)	1	1			X		$\frac{p+q}{p-q} \cdot \frac{p-q}{p+q} = 1$	Uso de la simplificación en la división de fracciones.
			c)	$\frac{(p-q)^2}{(p+q)^2}$	2		X			$\frac{p+q}{p-q} : \frac{p-q}{p+q} = \frac{(p-q)(p-q)}{(p+q)(p+q)}$	Confusión en el procedimiento de división de dos fracciones.
			d)	$\frac{(p+q)^2}{(p-q)^2}$	3	X					Respuesta Correcta
			e)	$\frac{2p+2q}{2p-2q}$	0				X	$\frac{p+q}{p-q} : \frac{p-q}{p+q} = \frac{p+q}{p-q} \cdot \frac{p+q}{p-q} = \frac{2p+2q}{2p-2q}$	Extensión de la suma algebraica a la multiplicación de fracciones.

4.3. Definición de la Muestra y los participantes

4.3.1. Población

La población corresponde a todos aquellos estudiantes secundarios o egresados, que participaron de forma activa en el Preuniversitario Víctor Jara durante el año 2019, en alguna de sus 7 sedes distribuídas por la Región Metropolitana.

Una de las características de los estudiantes son la variabilidad en la preparación previa que podrían haber tenido, ya que pudo haber sido bajo distintos tipos de formación académica, estrategias de enseñanza, niveles educativos o incluso la edad.

4.3.2. Muestra

Se realizó un muestreo de tipo probabilístico en el que cualquiera de los estudiantes del preuniversitario tenía la oportunidad de participar del estudio, dada la disponibilidad del horario de clases en el que se encontraba que era gestionado por la coordinadora del Preuniversitario.

Este procedimiento llevó a aplicar el test en un número reducido de cursos pertenecientes a 3 sedes del preuniversitario, dado que en el tiempo en que se comenzó a aplicar el test sucedieron acontecimientos en el país que redujeron considerablemente la asistencia de los estudiantes, por lo que tuvo que aplicarse en base a la disponibilidad de los profesores y de las clases, llegando a ser aplicado en 3 comunas diferentes donde el preuniversitario tenía dichas sedes.

La muestra finalmente, estuvo compuesta por estudiantes secundarios en su último año en su mayoría, donde 66 de ellos eran mujeres y 39 hombres, de los cuales cerca del 95 % tenía entre 15 y 23 años de edad.

Respecto al protocolo de aplicación, fue importante informarles que el test no medía su puntaje en la PSU sino su nivel de habilidad en el área de álgebra lineal para contar con su participación. Además, se realizó un permiso en doble hoja el

que incluía la autorización de un mayor de edad por el otro lado en caso de que el estudiante fuera menor de 18 años.

4.4. Criterios de Calidad

4.4.1. Validez de Contenido

La selección del Dominio Disciplinar se realizó bajo el criterio de que fuera el que mejor sirviera para los fines de la investigación, lo que terminó en la decisión de que el álgebra de 1er y 2do año de la enseñanza media era el dominio que mejor se ajustaba a la exploración de los diferentes tipos de errores seleccionados para esta investigación, dada la existencia habitual de errores asociados al cálculo, a la confusión de propiedades y la invención de las mismas.

Mediante pruebas inter-jueces realizadas a los ítems, se comprobó en primer lugar la correspondencia que existía entre el contenido evaluado y el tipo de situación problemática presentada en el ítem, de modo que ésta fuera relevante para cada uno de los subdominios escogidos, considerando contexto, enunciado y justificación. Por otra parte, se aseguraban de que los ítems fueran representativos del contenido disciplinar en su conjunto. En segundo lugar, comprobaron que los tipos de errores asociados a cada uno de los distractores se correspondiera con la definición asociada a sus tipologías (Concatenación, Yuxtaposición o No Epistemológico), de modo que de cada ítem pudiera extraerse información interpretable de forma válida posteriormente.

Luego de eso, se les mostraron los ítems validados a cuatro profesores del Preuniversitario a través de su coordinadora para que pudieran elegir entre los que consideraban más adecuados de acuerdo a la realidad de los estudiantes. De estos, los que tenían que ver con expresiones, variables o igualdades los consideraron adecuados, tanto en su nivel de dificultad como en su contenido, pero varios ítems de funciones fueron descartados debido a que los estudiantes del preuniversitario aún no habían visto ese contenido. Finalmente, 3 ítems de Resolución de Problemas fueron descartados porque consideraron que el nivel de dificultad de los mismos no estaba de acuerdo a lo que los estudiantes habían desarrollado hasta ese momento. De esta manera, el número de ítems por cada subtema quedó como se muestra en el Cuadro 4.2.

4.5. El Modelo de Crédito Parcial

El Modelo de Crédito Parcial (PCM) es un modelo unidimensional para el análisis de instrumentos de medición con dos o más categorías ordenadas de respuesta, pertenece a la familia de modelos de Rasch y es el más simple de los modelos de

Cuadro 4.2: Número de ítems aprobados por etapa

Subdimensión	Construídos	Aprobados	Elegidos
Expresiones	7	5	4
Variables	9	6	5
Igualdades	12	11	11
Funciones	15	12	5
Resolución	12	8	5
Total	54	42	30

respuesta al ítem para categorías ordenadas (Linden y Hambleton, 1997; p. 101). Dentro de sus aplicaciones se encuentran la evaluación de las funciones del lenguaje, ratings en el desempeño de infantes, problemas simulados de pacientes computarizados, ejemplos de ratings de escritura, medición del pensamiento crítico, medición de la comprensión de la ciencia, aplicaciones de la taxonomía de Solo, diagnóstico de errores en matemática (Adams, 1988) entre otros. Esta última aplicación es de la que se ocupará esta investigación: la medición del error a través del Modelo de Crédito Parcial.

Lo primero que realiza este modelo, es asumir que el respondiente posee cierta "cantidad" del constructo que se quiere medir, y que esa es la causa de las respuestas observadas al aplicar el instrumento (Wilson, 2005, p. 10). Para ello, asumiremos que nuestro constructo corresponde a la habilidad latente que posee el estudiante respecto del tema que se le está preguntando: en caso de responder correctamente, se le asignará el máximo puntaje correspondiente a 3, que a su vez, significa que no posee o realizó ningún tipo de error. Si el estudiante comete el nivel más simple de error (No Epistemológico), se le asignará el puntaje 2, considerando que su error pudo haber sido accidental, o del tipo descrito anteriormente. Si selecciona la opción que corresponde a un error moderado (Yuxtaposición), se le asignará el puntaje de 1. Finalmente, si selecciona el peor tipo de error (Concatenación), se le asignarán 0 puntos debido a que no posee suficiente habilidad latente respecto al tema que se trata esa problemática como para tener algún puntaje. Pese a que nuestro interés está en el error, con esta asignación de puntajes es más plausible una correcta interpretación de los resultados debido a que entre mayor sea la habilidad latente, mayor será el puntaje que se espera que obtenga el estudiante.

Así, las puntuaciones quedarían como se indica en el Cuadro 4.3. En este caso, el uso del modelo de Crédito Parcial permite diferenciar tipos de errores no sistemáticos mediante la asignación de diferentes puntajes. De esta forma, el análisis que se presentará a continuación se centrará, principalmente, en el análisis de dichos errores.

Cuadro 4.3: Puntuación de los tipos de error.

Puntaje	Tipo de Error	Forma Simple
0	Concatenación	Complejo
1	Yuxtaposición	Moderado
2	No Epistemológico	Simple
3	Ninguno	Sin error

El análisis de los datos se realizó a través del software de modelamiento generalizado de Respuesta al Ítem CONQUEST 2.0 ©(2017), el cual brindó estimaciones de los parámetros de dificultad de los ítems y de las alternativas, a través de tablas, mapas, gráficos de las curvas características y de la probabilidad acumulada.

A partir de las tablas de estimación de los parámetros se realizaron análisis de los ítems con el fin de clasificarlos de acuerdo a su valor de dificultad θ y relacionándolos con el tema específico al que correspondían, como sugieren Linden y Hambleton (1997; p. 109). Así mismo, se verificó el ajuste de estos valores a las características del instrumento mediante los Unweighted Mean Square, contrastando estos valores de modo que no correspondieran a casos aislados con los datos entregados por el Weighted Mean Square (Masters y Right, 1982; p. 100), lo que permitió un diagnóstico general del instrumento sobre la "bondad del ajuste" de sus ítems. Adicionalmente a eso, se determinaron los índices de discriminación a partir de las estadísticas tradicionales de los ítems otorgadas por el mismo software, los que se analizaron de manera comparativa entre los ítems considerando las sugerencias de Andrich y Marais (2019; p. 153) de que a mayor valor de discriminación del ítem es un indicio de buen funcionamiento, siempre y cuando no sea demasiado alta ya que se pierde dependencia local, y que una baja discriminación podría ser considerado como una falta de ajuste del ítem.

En cuanto al nivel de habilidad de los ítems, se realizó posteriormente un estudio a partir de los valores de dificultad θ, donde la probabilidad de contestar la alternativa con puntaje i o superior era del 50 %, llamados Thresholds o "τ". Estos valores θ, se visualizaron en los gráficos llamados "Cumulative Probability Curve(s)" (Curvas de Probabilidad Acumulada) mostrados en la siguiente sección, donde a pesar de haber 4 puntajes diferentes (0, 1, 2 y 3), se consideran sólo 3 Thresholds (asociados a los puntajes 1, 2 y 3), debido a que la probabilidad de obtener puntaje 0 o superior es siempre del 100 %. Es por ello que se consideran los valores τ_i de la siguiente forma (Andrich y Marais, 2019; p. 235):

- τ_1: Corresponde al valor de θ, para el cual el estudiante tiene un 50 % de probabilidad de contestar las alternativas con puntaje 1 o superior.

- τ_2: Corresponde al valor de θ, para el cual el estudiante tiene un 50 % de

probabilidad de contestar las alternativas con puntaje 2 o superior.

- τ_3: Corresponde al valor de θ, para el cual el estudiante tiene un 50 % de probabilidad de contestar la alternativa con puntaje 3.

Esto permitió representar de forma gráfica la dificultad de cada uno de los ítems considerando sus 3 Thresholds, y clasificarlas posteriormente en cuatro categorías de acuerdo a su distribución.

Posteriormente, se analizó el funcionamiento de los puntajes en cada uno de los ítems, a través de su comportamiento en los gráficos llamados "Characteristic Curve(s) By Score" (Curva Característica por Puntuación), donde se observa la probabilidad de cometer un determinado tipo de error en función de la habilidad latente del estudiante mencionado como "Latent Trait" o Logits en el gráfico a continuación:

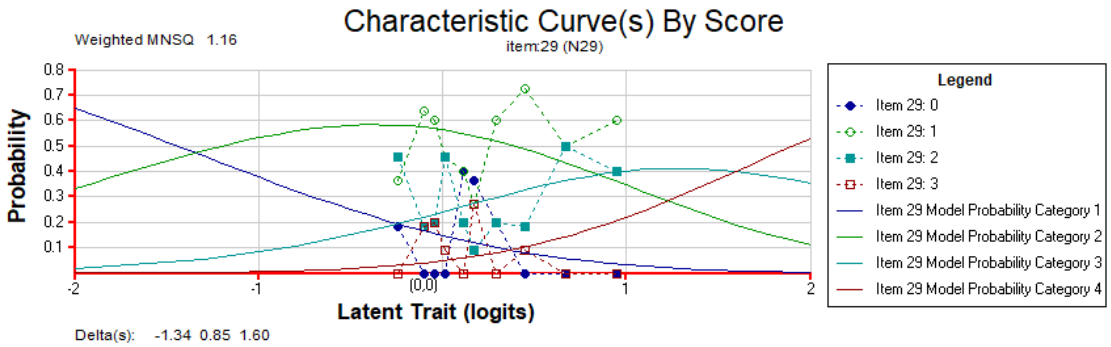


Figura 4.2: Funcionamiento de los puntajes

Este análisis se centró principalmente en los valores Delta δ_i de estos gráficos, que corresponden a los puntos de intersección entre dos curvas características de puntuaciones consecutivas, dicho de otra manera, en los valores de habilidad latente donde la probabilidad de cometer un determinado nivel de error o el siguiente, es la misma. Es así como podrían definirse cada uno de los δ_i de la siguiente forma:

- δ_1: Corresponde al valor de habilidad latente donde la probabilidad de cometer un error de Concatenación o de Yuxtaposición, es la misma.

- δ_2: Corresponde al valor de habilidad latente donde la probabilidad de cometer un error de Yuxtaposición o No Epistemológico, es la misma.

- δ_3: Corresponde al valor de habilidad latente donde la probabilidad de cometer un error No Epistemológico o de responder correctamente, es la misma.

De estos valores se analizó su orden (de menor a mayor), tendencia (todos hacia la izquierda, hacia el centro o hacia la derecha), y también se categorizaron entre los

que poseían una o dos de sus curvas con un valor de probabilidad menor a las otras curvas para cualquier valor de habilidad.

Capítulo 5

Análisis de Resultados

Con los primeros resultados de la aplicación del test se obtubieron algunos valores que permitieron evaluar el comportamiento general del test y de su puntuación:

Cuadro 5.1: Estadísticos Generales del Instrumento

Estadístico	**Valor**
Tamaño de Muestra	105
Error de Medición	5,76
Coeficiente Alpha	0,78

Estos valores indican que de un total de 90 puntos máximo que poseía el test, el puntaje de cada estudiante podía tener un margen de error de 5 puntos aproximadamente, y además, que con una fiabilidad dada por un alpha de Cronbach de $\alpha = 0{,}78$ los resultados del test tienen una buena relación entre sí con el constructo que se quería medir (Gempp, 2006), ya que es un valor lo suficientemente cercano a 1 dado el número de personas que rindieron el test.

5.1. Funcionamiento de los ítems

5.1.1. Parámetros estimados por ítem

Se optó por los parámetros estimados de los ítems y no de las personas, debido a que como se mencionaba anteriormente, los profesores buscan una forma de ir mejorando sus diagnósticos, lo que puede realizarse mediante el mejoramiento del instrumento que permita identificar deficiencias o errores particulares de rendimiento de manera rápida y útil para los docentes.

Siendo así, los parámetros estimados para los niveles de dificultad de cada uno de los ítems, se calcularon mediante la “Probabilidad Máxima Conjunta” o “Joint

Maximum Likelihood" (JML) como lo mencionan Linden y Hambleton (1997, p. 109), mediante la siguiente expresión a través del software Conquest:

$$\hat{\delta}_{ih}^{(t+1)} = \hat{\delta}_{ih}^{(t)} - \frac{-S_{ih} + \sum_{r}^{M-1} N_r \sum_{k=h}^{m_i} P_{irk}^{(t)}}{-\sum_{r}^{M-1} N_r \left[\sum_{k=h}^{m_i} P_{irk}^{(t)} - \left(\sum_{k=h}^{m_i} P_{irk}^{(t)} \right)^2 \right]} \tag{5.1}$$

Donde $\hat{\delta}_{ih}^{(t)}$ es el estimado de δ_{ih} luego de t iteraciones, N_r es el número de personas con puntaje r, y M es el puntaje máximo del instrumento. De esta manera, las estimaciones fluctuaron en el rango de [-0.57;0.638], de donde se observan tanto en la Tabla A como en el Mapa A de los "Model Parameters Estimates" que hay cuatro ítems muy por debajo del resto del grupo, que son los ítems 8 (-0.404), 9 (-0.445), 20 (-0.57) y 25 (-0.479) que pertenecen a Igualdades, Variables y Expresiones. Por el otro lado, se observaron dos ítems superiores al resto del grupo, los ítems 19 (0.638) y el 22 (0.604), que pertenecen a Funciones y Expresiones, que son a su vez los que requieren un mayor nivel de habilidad:

Cuadro 5.2: Ejemplos de parámetros con valores extremos

Valores Extremos Bajos		Valores Extremos Altos	
N° Ítem	Estimate	N° Ítem	Estimate
8	-0.404	19	0.638
9	-0.445	22	0.604
20	-0.570		
25	-0.479		

5.1.2. Ajuste de los ítems

Para la estimación de lo adaptado que estaba cada uno de los ítems a la medición del constructo, se calculó la diferencia estandarizada (z_{ij}) entre la respuesta esperada y la observada de cada ítem "i" para cada persona "j". Esto se realizó a través de la ecuación (Linden y Hambleton, 1997, p. 113):

$$z_{ij} = \frac{x_{ij} - E_{ij}}{\sqrt{W_{ij}}} \tag{5.2}$$

Donde x_{ij} es la respuesta observada, E_{ij} es la respuesta esperada y W_{ij} la varianza de x_{ij}, los que conforman la diferencia estandarizada.

Unweighted Mean Square Fit (Outfit)

De las diferencias obtenidas anteriormente para cada persona, entre el dato real y el esperado, se calculó un promedio de éstas para cada uno de los ítems, para poder obtener el Unweighted Mean Square, al que se le denominó u_i (Masters y Right, 1982, p.100):

$$u_i = \sum_{j}^{N} \frac{z_{ij}^2}{N} \tag{5.3}$$

Por otro lado, Wu, Tam y Yen (2016, p. 150) mencionan que la mayoría de los ítems pueden tener valores fit-t (que se muestran en la columna de la derecha en la siguiente tabla) entre -2 y +2 con un nivel de confianza del 95 %, colocando especial atención a los que poseen un valor $|x| > 3$. De esta manera, las estimaciones para los fit-t de Unweighted Mean Square (Tabla B) fluctuaron en el rango de $[-2{,}1; 4{,}1]$, de los cuales se destacan 3 ítems particularmente altos, el N° 8 (t = 1,9), el N° 5 (t = 2,2) y el N° 22 (t = 4,1), lo que significa que el N°8 presenta respuestas altas pero aún dentro de los límites esperados, el N° 5 muestra sólo indicios de respuestas más altas de lo esperado para la alta dificultad del ítem, pero el N° 22, presenta un fit-t tan alto, que podría significar que existen respuestas con puntaje inesperadamente alto (3 o 4) para este ítem particularmente difícil, de personas con habilidad más bien baja.

En el otro extremo, existen tres ítems con un fit-t particularmente bajo que son el N° 4 ($t = -1{,}9$), el N° 2 ($t = -1{,}9$) y el N° 27 ($t = -2{,}1$), lo que puede interpretarse como que el N° 4 y el N° 2 están aún dentro de los límites de respuesta esperados, pero el N° 27 posee indicios de ser un ítem que requería de una habilidad más baja de la estimada para el grupo de estudiantes.

Cuadro 5.3: Ejemplos de valores extremos para Outfit

N°	UNWEIGHTED FIT		
ITEM	MNSQ	CI	T
27	0.73	(0.73 ; 1.27)	-2.1
2	0.75	(0.73 ; 1.27)	-1.9
4	0.75	(0.73 ; 1.27)	-1.9
8	1.28	(0.73 ; 1.27)	1.9
5	1.32	(0.73 ; 1.27)	2.2
22	1.68	(0.73 ; 1.27)	4.1

Weighted Mean Square Fit (Infit)

Para evitar un error en el análisis como sucede con los "outliers" del Unweighted Mean Square Fit debido a la posibilidad de casos aislados, se utilizaron las mismas

diferencias estandarizadas z_{ij}, de modo que ponderara el peso de cada una de las varianzas de los ítems, a través de la siguiente ecuación (Masters y Right, 1982, p.100):

$$v_i = \frac{\sum_{j}^{N} W_{ij} z_{ij}^2}{\sum_{j}^{N} W_{ij}} \tag{5.4}$$

Los valores MNSQ de los Weighted Mean Square Fit (Tabla B) fluctuaron en el rango de $[0,76; 1,59]$, de los cuales sólo uno de ellos tuvo un valor superior al adecuado para la productividad de la medición ($[0,5; 1,5]$), correspondiente al ítem N° 22, como se muestra en la siguiente tabla de los 3 ítems con valores MNSQ más altos:

Cuadro 5.4: Ejemplos de valores extremos para Infit

N°	WEIGHTED FIT		
ITEM	MNSQ	CI	T
13	1.20	(0.83 ; 1.17)	2.2
5	1.24	(0.83 ; 1.17)	2.6
22	1.59	(0.80 ; 1.20)	5.1

Esto quiere decir, que el ítem es en cierto grado improductivo para la medición, pero no degradante (Rasch; 2002), y dado que es un valor ligeramente superior a los límites, puede considerarse como evidencia para investigar posteriormente con mayor profundidad los datos asociados al ítem 22.

Por otro lado, todos los demás ítems obtuvieron valores de MNSQ superiores al mínimo recomendable para la productividad de la medición $(0,5)$, lo que sugiere que los ítems se ajustan al modelo con el que fueron evaluados. A continuación, se muestran los 3 ítems con los valores MNSQ más bajos:

Cuadro 5.5: Ejemplos de valores extremos para Infit

N°	WEIGHTED FIT		
ITEM	MNSQ	CI	T
27	0.76	(0.81 ; 1.19)	-2.7
2	0.80	(0.84 ; 1.16)	-2.7
28	0.86	(0.83 ; 1.17)	-1.7

Desde ambas miradas, es decir, tanto desde el Unweighted Mean Square Fit como desde el Weighted Mean Square Fit, se observa que particularmente los ítems N° 22 y N° 27 necesitan una evaluación, ajuste, reemplazo o eliminación del test (Masters y Right, 1982, p. 102).

Síntesis del Comportamiento de los ítems

La mayoría de los ítems tuvieron un comportamiento dentro de los valores esperados, sin embargo, existen valores estimados de θ, unweighted y weighted, que muestran indicios de que son ítems que requieren de una revisión de su enunciado, contexto o alternativas, como los del Cuadro 5.6.

Cuadro 5.6: Síntesis de valores extremos de los ítems

N° Ítem	Estimates	Unweighted	Weighted
2		-1.9	-2.7
4		-1.9	
5		2.2	2.6
8	-0.404	1.9	
9	-0.445		
13			2.2
19	0.638		
20	-0.570		
22	0.604	4.1	5.1
25	-0.479		
27		-2.1	-2.7
28			-1.7

De los ítems con valores extremos, existen algunos que acumulan hasta tres valores extremos inclusive (como el ítem N° 22), mostrando evidencia de que el ítem está aportando menos para la medición del constructo.

5.1.3. Clasificación de ítems por Thresholds en los CCC

Los Thresholds (τ_i) como se mencionó en el capítulo anterior, corresponden a los valores de dificultad θ, para el cual el estudiante tiene un 50 % de probabilidad de contestar las alternativas con puntaje i o superior, que son posibles de visualizar en las Curvas de Probabilidad Acumulada (CCC).

Para obtener el valor de θ para el cual su probabilidad es 0,5, se utilizó la siguiente ecuación:

$$P_{ijx} = \frac{e^{\sum_{k=0}^{x}(\theta_j - \delta_{ik})}}{\sum_{h=0}^{m_j} e^{\sum_{k=0}^{x}(\theta_j - \delta_{ik})}}, x = 0, 1, ..., m_i \tag{5.5}$$

En el Gráfico 1, se muestran los Thresholds para cada ítem, de los que se construyeron 4 clasificaciones, basándose en sus distribuciones hacia la parte positiva, negativa o central. Los ejemplos para los Thresholds encontrados en la literatura, usualmente presentan límites entre los valores -1 y +1, por lo que se utilizaron estos

valores como referencia para clasificar a los ítems, en función de los que tenían la distribución de sus Threshold más orientados hacia la parte positiva, hacia la negativa, o con valores extremos superiores al valor absoluto de 1, como se detalla a continuación:

Fáciles

Se consideraron como fáciles los ítems con una distribución de sus Thresholds hacia la parte negativa, con el menor de sus Threshold cercano o menor a -1, y su mayor Threshold cercano o menor a 0. Estos ítems corresponden al N° 8, 9, 20 y 25 (sus valores pueden consultarse en la Tabla C "Traditional Item Statistics y visualizarse en el Gráfico 1 "Calibración de los ítems").

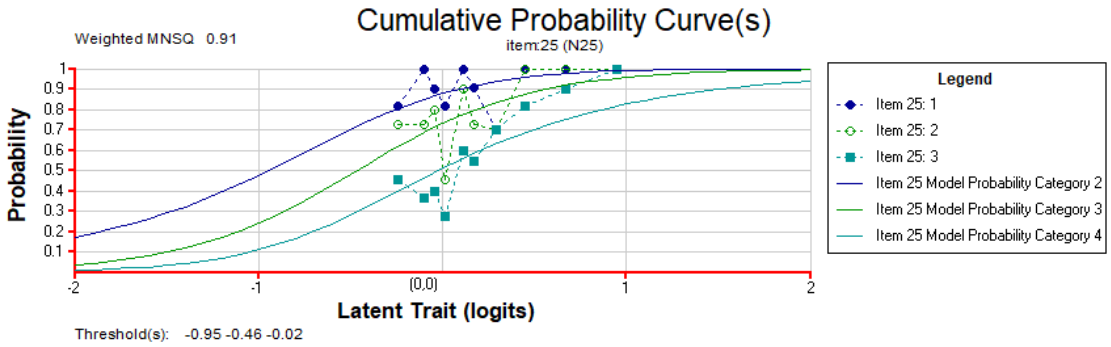


Figura 5.1: Ítem Fácil por Thresholds

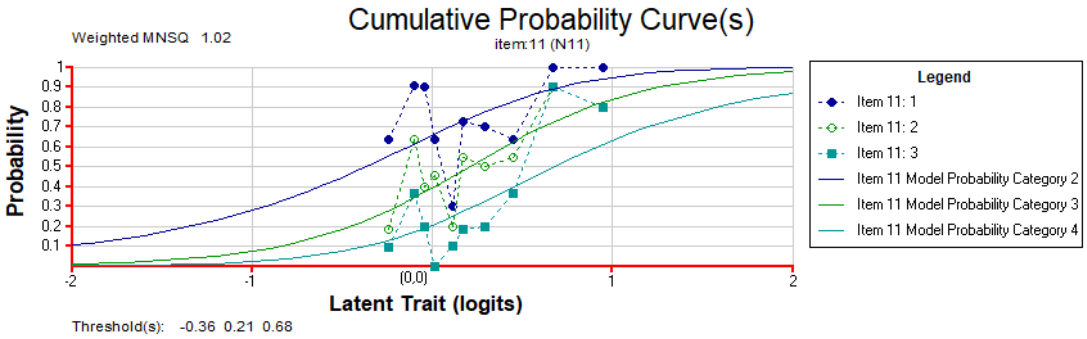
Estándar

Se consideraron como Estándar los ítems cuya distribución de sus Thresholds están aproximadamente dentro del intervalo [-1,+1]. Estos ítems corresponden al N° 2, 3, 4, 5, 6, 10, 11, 12, 14, 15, 21, 22, 23, 27 y el 30. Se consideraron además los ítems que poseen uno de sus Thresholds fuera de ese límite pero el otro pertenece aún a él, como son el N° 7, 17, 18, 24, 26 y 28. Cabe destacar, que en todos estos ítems se visualiza un nivel de respuesta de las alternativas con puntaje 1 y 2 entre el 40 y 60 % y permite brindarle al instrumento una mejor distribución de la dificultad de sus ítems.

Difíciles

Los ítems Difíciles se consideraron como los que poseen una distribución de sus Thresholds mayormente hacia la parte positiva, con alguno de sus Thresholds que supera el límite +1. Estos corresponden al N° 13, 16, 18, 19 y 29. Además, los que poseen su menor Threshold relativamente alto comparado con los demás, es decir

Figura 5.2: Curva de Probabilidad Acumulada N°11



cercano o superior a 0, corresponden a los ítems N° 13 y 19. El ítem N° 22 se consideró dentro de los ítems Difíciles debido a que aunque a pesar de que su distribución se mantiene dentro de los intervalos, esta es sólo positiva. El N° 18 mantiene su Threshold inferior superior a -1 lo que lo hace sensible al resto de probabilidades igual como los ítems Estándar.

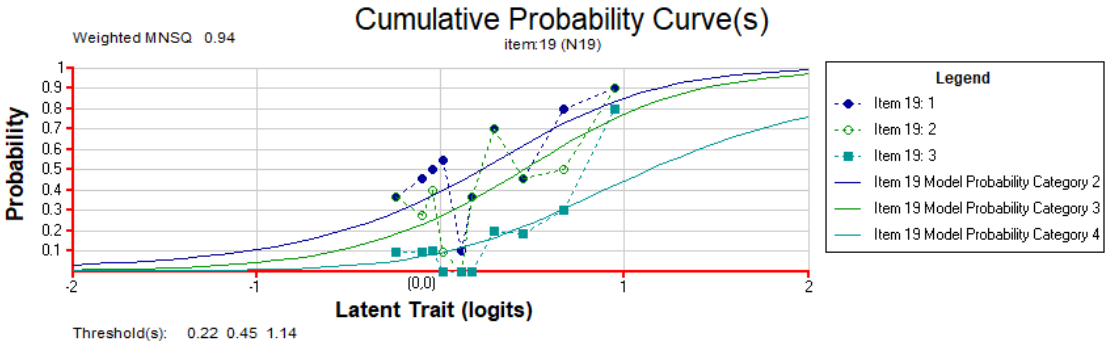


Figura 5.3: Ítem Difícil por Thresholds

Ambiguos

Estos ítems corresponden a los que poseen sus Thresholds máximos fuera del intervalo establecido, lo que no indica que no sean un aporte para la medición del constructo, pero si que se vuelve necesario darles una mirada exhaustiva al comportamiento de sus distractores, ya que en ambos casos muestran una tendencia hacia el distractor de 1 punto, incluso del 50 %. Corresponden al N° 16 y al 29. En el siguiente gráfico se muestra un ejemplo de 4 ítems, en los que se aprecian las clasificaciones de uno Fácil (N° 25), Estándar (N° 12), Difícil (N° 19) y otro Ambiguo (N° 16), donde los Thresholds están marcados con un diamante negro, similar al mostrado por Wu, Tam Y Yen (2016, p. 172):

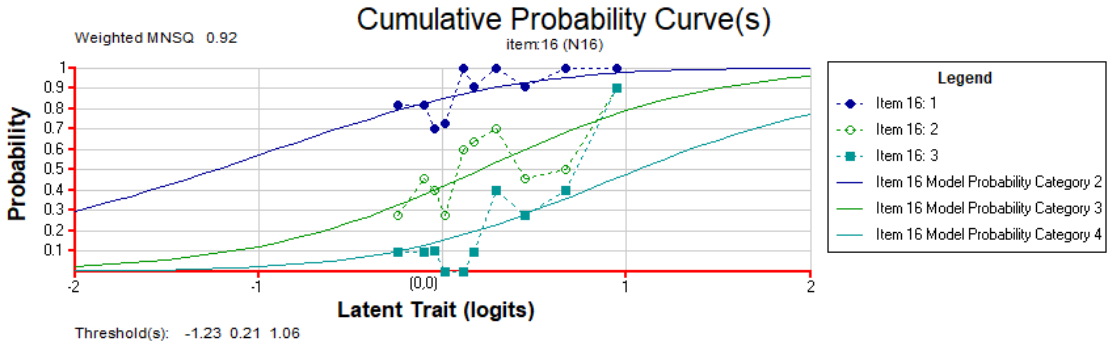


Figura 5.4: Ítem Ambiguo por Thresholds

Distribución de Dificultad de los ítems

Considerando el promedio de los Thresholds para cada uno de los ítems, fue posible obtener su valor de dificultad individual de manera que pudieran ser ordenados de forma creciente para facilidad de su lectura en el siguiente gráfico:

Figura 5.5: Nivel de dificultad ordenados de menor a mayor

5.2. Funcionamiento de los Distractores

Los δ se consideran como los valores θ donde se intersectan las curvas de probabilidad de categorías adyacentes (Masters y Right, 1982, p. 43), las que pueden apreciarse en las curvas características del ítem. Para su análisis en conjunto se realizó la Tabla G donde se fueron ordenando por separado de menor a mayor el δ_1, δ_2 y δ_3. Se consideró de referencia además los límites $[-1; 1]$, para clasificar dichas intersecciones en tres categorías, los Delta Inferiores (δ^-), los Delta Neutros (δ^0) y los Delta Superiores (δ^+).

5.2.1. Comportamiento Esperado

El comportamiento esperado de los distractores es que aparezcan todos en orden, de menor a mayor, lo que revelaría a su vez que cada uno de ellos tuvo una probabi-

lidad mayor a los demás para ser elegido en un intervalo de habilidad en particular, como son los ítems N° 16, 24, 26 y 29. A continuación se presenta de ejemplo, la Curva Característica del Ítem N° 29:

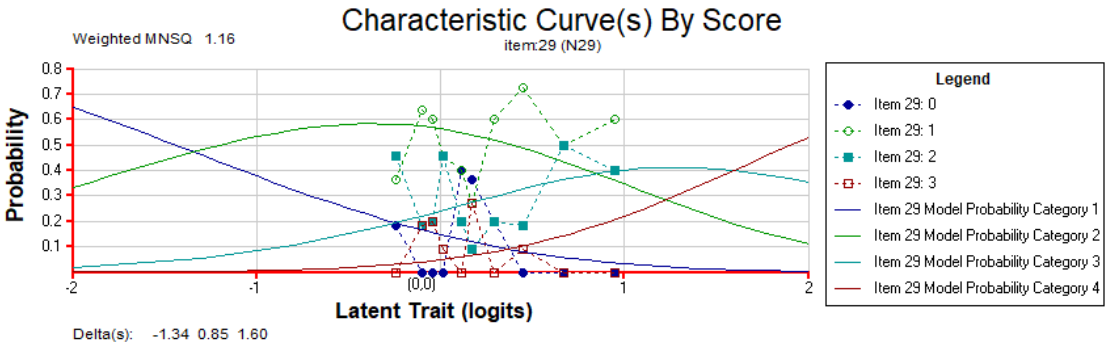


Figura 5.6: Comportamiento esperado de Distractores

En las curvas características del ítem 29, se aprecia que la probabilidad de cometer un error de Concatenación o de Yuxtaposición, es la misma para un $\delta_1 = -1{,}34$, por lo que los estudiantes que poseen habilidad en el intervalo $[-2; -1{,}34]$ tienen mayor probabilidad de cometer un error de Concatenación. Luego, los estudiantes que poseen habilidad en el intervalo $[-1{,}34; 0{,}85]$ tienen mayor probabilidad de cometer un error de Yuxtaposición, los que poseen habilidad en el intervalo $[0{,}85; 1{,}60]$ tienen mayor probabilidad de cometer un error de tipo No Epistemológico y los que poseen un nivel de habilidad en el intervalo $[1{,}60; 2]$ tienen mayor probabilidad de responder correctamente. Con estos datos, se observa que todas las categorías tuvieron su intervalo de protagonismo, sin embargo, el error de Yuxtaposición (puntaje 1) fue el que tuvo mayor presencia para niveles de habilidad que iban desde $-1,34$ hasta $0,85$, lo que muestra una evidencia de que este tipo de error es el que predominaba en los estudiantes.

5.2.2. Ítems con una Categoría de bajo interés

Existen algunos ítems que poseen una de sus curvas características mas baja que cualquiera de las otras curvas. En estos casos puede estar sucediendo que ningún estudiante esté de hecho cometiendo ese tipo de error, o bien, que sea necesario realizar una evaluación de ese distractor.

A continuación, se mostrarán tres ejemplos de ítems con categorías de bajo interés, correspondientes a un ítem categorizado anteriormente como fácil, otro estándar y otro difícil:

Ítem fácil con categoría de bajo interés

El valor estimado para el ítem 9 es de $-0{,}445$, y como puede apreciarse todas sus curvas están más presentes en el lado izquierdo del eje de habilidad θ lo que muestra evidencia de que era un ítem mas bien fácil. Los valores donde se cruzan sus curvas (δ) comienzan en $\delta_1 = -1{,}10$, es decir, un estudiante que posea hasta ese nivel de habilidad tiene mayor probabilidad de cometer un error de Concatenación (puntaje 0). Sin embargo, observamos que $\delta_2 = 1{,}79$ está bastante más alejado de las otras curvas, y es porque la curva asociada al error de Yuxtaposición (puntaje 2) se mantiene baja a lo largo de todo el recorrido de habilidad latente, de modo que prácticamente no tiene presencia sugiriendo que estas alternativas no fueron buenos distractores, el valor que nos sería útil es el punto de intersección entre las curvas asociadas al error de Yuxtaposición y a la respuesta correcta (puntajes 1 y 3 respectivamente), que es el valor a partir del cual cualquier estudiante que posea ese nivel de habilidad o superior (aproximadamente $-0{,}2$) tendrá cerca de un 40 % o más probabilidad de contestar la alternativa correcta.

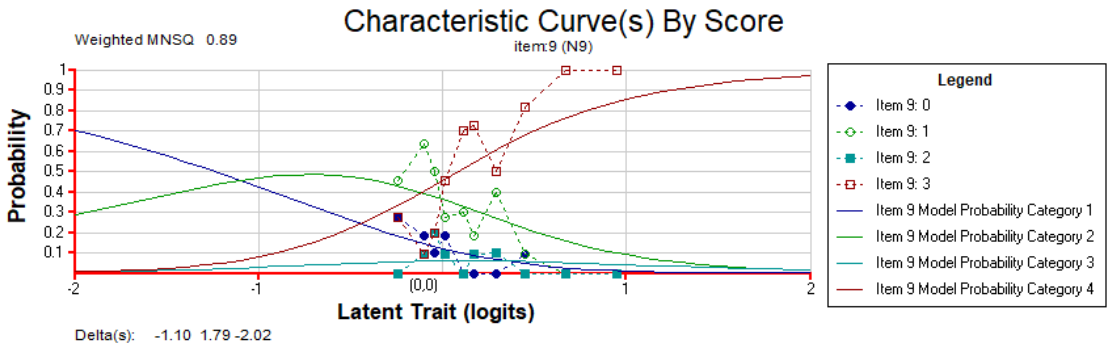


Figura 5.7: ICC de Ítem Fácil con categoría de bajo interés

Ítem Estándar con categoría de bajo interés

El ítem estándar escogido como ejemplo es el N° 30, cuyo valor estimado es de $-0{,}129$ lo que lo deja más bien centrado en el gráfico como puede apreciarse en la figura a continuación. Observamos que la curva asociada al error de Yuxtaposición (ptje. 1, de color verde) jamás sobrepasa a las otras curvas, por lo que nunca es la más probable para ningún valor de habilidad. El error de Concatenación (ptje. 0) es el más probable para valores inferiores a $\theta = -0{,}5$ que es donde se intersecta con la curva asociada al error de tipo No Epistemológico (ptje. 2), poco antes del δ_1. Los estudiantes que poseen un nivel de habilidad en el intervalo $\theta \in [-0{,}5; 0{,}55]$ este último valor correspondiente al δ_3 poseen mayor probabilidad de contestar la alternativa asociada al error de tipo No Epistemológico (ptje. 2). Luego de eso los

que poseen mayor habilidad ya tienen desde un 40 % de probabilidad de contestar correctamente:

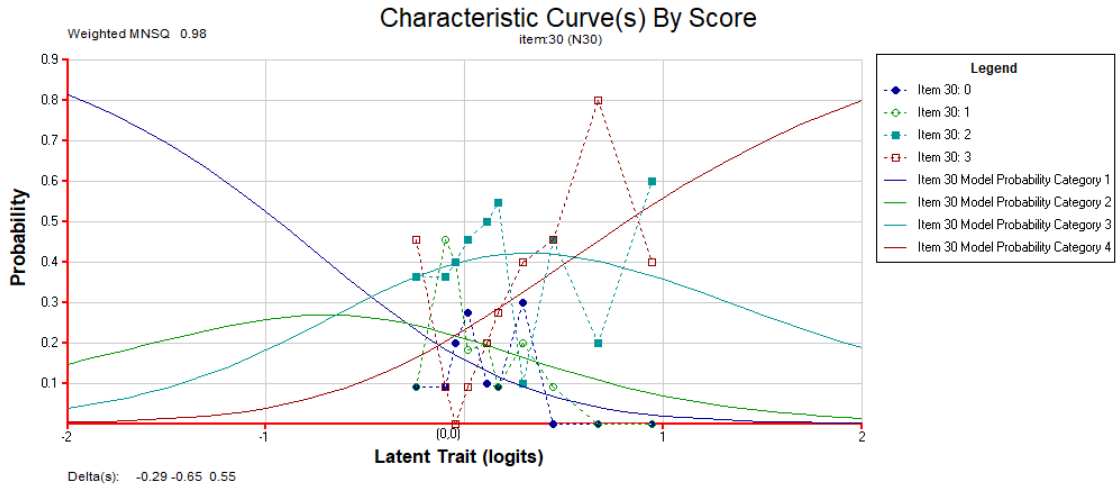


Figura 5.8: ICC de Ítem Estándar con categoría de bajo interés

Ítem Difícil con categoría de bajo interés

Uno de los ítems difíciles escogidos como ejemplo fue el 13 con un valor estimado de 0,389 que se perciben en la concentración de las curvas hacia la derecha. Este también posee sólo tres de las cuatro curvas características, predominando a lo largo de todo el eje de habilidad θ. En este caso el error de Yuxtaposición (ptje. 1) fue el que tenía menor probabilidad de ser cometido por parte de los estudiantes, y se puede apreciar que existe una mayor probabilidad de escoger el error de Concatenación (ptje. 0) cuando se tiene $\theta < 0{,}2$ que es donde aproximadamente se intersecta con la curva asociada al error de tipo No Epistemológico (ptje. 2), y una mayor probabilidad de cometer un error de tipo No Epistemológico para los que posean $\theta \in [0{,}2; 0{,}68]$. Luego de eso, para valores superiores de habilidad hay una mayor probabilidad de que el estudiante conteste correctamente:

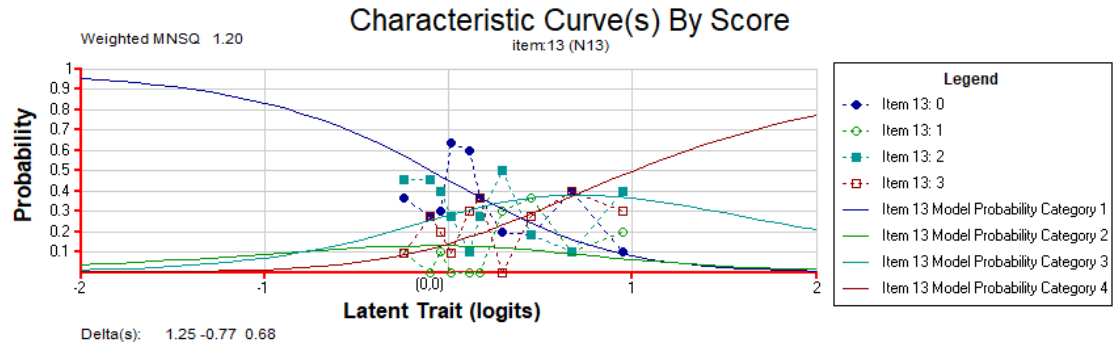


Figura 5.9: ICC de Ítem Difícil con categoría de bajo interés

Como se puede apreciar en el gráfico, la curva asociada al error de Yuxtaposición (ptje. 1) no sobrepasa para ningún valor de habilidad a las demás curvas, indicando que los estudiantes no suelen cometer este tipo de error en el ítem N° 13.

5.2.3. Ítems con dos Categorías de bajo interés

Hubieron algunos ítems cuyas dos categorías centrales, los errores de Yuxtaposición y de tipo No Epistemológico (ptjes. 1 y 2), se mantuvieron por debajo del nivel de probabilidad de las curvas asociadas al error de Concatenación y de la respuesta correcta (ptjes. 0 y 3) para todos los valores de habilidad del ítem, como se observa en la siguiente figura:

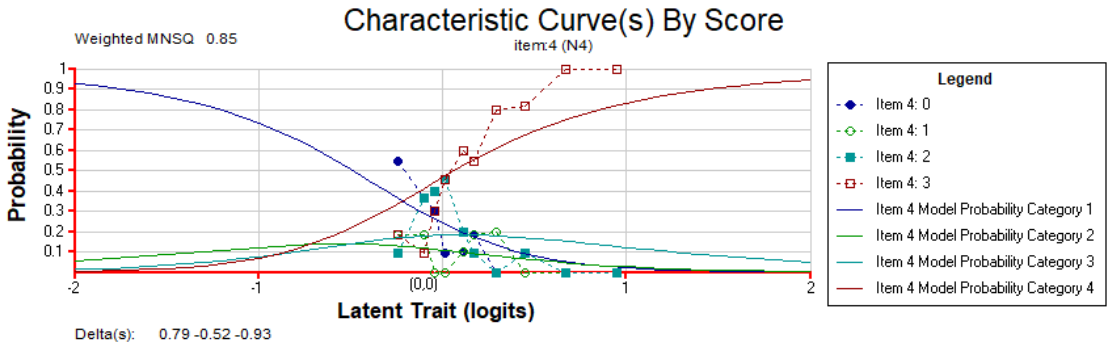


Figura 5.10: ICC de Ítem con dos categorías de bajo interés.

En este caso, sólo existen dos curvas que predominan para los distintos niveles de habilidad, la curva asociada a los errores de Concatenación (ptje. 0) que predomina hasta valores de $\theta = -0{,}2$ aproximadamente, a partir de donde entonces la curva asociada a la respuesta correcta (ptje. 3) predomina hasta el final. Los otros δ_i no se vuelven oportunos debido a la poca elección de los estudiantes sobre estas alternativas.

Esto podría sugerir, que los estudiantes están cometiendo sólo un tipo de error en este ítem (el de Concatenación), o bien, que los distractores asociados a los errores de Yuxtaposición y No Epistemológicos necesitan mejorarse o fusionarse en uno solo (Linden y Hambleton, 1997; p.104).

5.3. Relación entre Dificultad y Distractores

Al correlacionar el porcentaje de personas que contestó cada uno de los tipos de error con la dificultad de los ítemes (Anexo F), se observa una correlación de $r_0 = 0{,}84$ con la probabilidad de observar el error de Concatenación, y de $r_3 = -0,78$ con la probabilidad de observar la alternativa correcta. Esto significa que a mayor dificultad del ítem existe una mayor probabilidad de observar un error de concatenación

(ptje. 0), y menor probabilidad de contestar correctamente (ptje. 3). Sin embargo, la correlación entre la dificultad estimada de los ítemes y la probabilidad de observar los errores de Yuxtaposición o No Epistomológico (ptj. 1 y 2), son cercanas a cero, es decir, no se observa una evidencia de tal relación.

Es por esto, que a continuación se utilizaron otras herramientas entregadas por CONQUEST para realizar las comparaciones con estos otros dos tipos de errores: las Curvas Características del Ítem (ICC) y los Thresholds (τ).

Error de Concatenación y Yuxtaposición

La relación entre estos dos tipos de errores se puede visualizar a través del Delta 1 (δ_1) de las ICC, ya que es el valor -en escala logit- donde la probabilidad de cometer el error tipo 0 y 1 es la misma (Wu et. al., 2016).

Al estimar la correlación entre los δ_1 con su respectivo valor de dificultad de cada ítem, se obtiene una correlación de $r_x = 0{,}64$. Esto sugiere que la dificultad del ítem y el punto donde ocurre la transición entre la probabilidad de observar un error tipo 0 y un error tipo 1, está relacionados. Esto quiere decir que la dificultad de un ítem, tiene cierta influencia en la probabilidad de cometer un error de concatenación o de yuxtaposición.

Error de Yuxtaposición y No Epistemológico

Para observar la relación entre estos dos tipos de errores, se utilizaron los Mapas de Distribución de Hailidad Latente y Thresholds (Figura B.2), que muestran el valor estimado para la transición entre dos niveles de dificultad adjuntos, en este caso el nivel 1 y el 2.

En este mapa, se aprecia que el valor de transición entre los errores de concatenación y yuxtaposición ($x.1$) están más abajo en la escala de dificultad que los valores de transición entre los errores de yuxtaposición y no epistemológicos ($x.2$). Esto quiere decir, que en general es más difícil cometer un error de tipo No epistemológico, por lo que a mayor dificultad del ítem, menor será la probabilidad de cometer un error No epistemológico, manteniendo la premisa de que la dificultad de un ítem, influye en la probabilidad de cometer determinado tipo de error.

Capítulo 6

Discusión de Resultados

Las distintas perspectivas analizadas acerca de la medición de un conocimiento o habilidad, coinciden en que una evaluación que tenga la intención de ayudar en la mejora de un aprendizaje, necesita brindarle de un insumo al estudiante que le permita al menos identificar el "vacío" que le falta por aprender (Brookhart, 2003), incluso cuando este vacío, pueda tratarse de un error. Esto llevó a diversos autores (como Socas, 2000; Tall, 1989; Radatz, 1979; entre otros) a anticiparse a los distintos tipos de errores que pueden ser cometidos frente a ciertas situaciones problemáticas, y catalogar tipologías que permitan construir problemas basándose en dichas respuestas (González, 2015). Sin embargo, no se ha llegado a un consenso sobre cual es la más adecuada, ya que esto depende también del punto de vista sobre el aprendizaje que se tome en consideración.

Así, tomando en cuenta los hallazgos de este proyecto, a continuación se discute el uso de una tipología de errores que podría ser de utilidad en los test de alternativas -ampliamente usados en este tiempo dada su facilidad para ser analizados-, y de un elemento influyente en su elaboración como lo es el nivel de dificultad.

6.1. Dificultad y Tipologías de Errores

Uno de los aspectos fuertemente vinculados con la medición del error, fue la graduación de la dificultad de cada situación problemática (Tall, 1989, citado en Castro, 2000, p. 136) y su uso para compararlas con el nivel de habilidad de los estudiantes.

Socas (citado en Rico, 2000) clasificó los errores de acuerdo a la forma en que se llega a una respuesta equivocada, sin embargo, este no consideró que el tipo de error cometido podría tener una relación con el nivel de dificultad de la situación problemática. Como por ejemplo, que a menor dificultad, era más probable cometer errores no epistemológicos (tipo 2), y a mayor dificultad, era más probable cometer

errores de concatenación (tipo 0).

Además, como los resultados del PCM mostraron los valores para los índices de dificultad y de habilidad de los estudiantes en una escala equivalente (Apéndice B.1.), estos serían comparables entre sí, y como Productos Notables tuvo en su mayoría ítems de baja dificultad, se podría llegar a asumir que en esta área se puede adoptar un enfoque pedagógico más orientado hacia resolver errores de tipo No Epistemológico. De manera contraria, sucedió con el área de Resolución de Problemas que tuvo en su mayoría ítems de alta dificultad.

6.2. Vacíos de Aprendizaje y Tipologías de Error

Desde un punto de vista más general, los resultados anteriores nos llevan a plantear que el nivel de habilidad necesario para cometer un error de tipo No epistemológico, debe ser mayor que para cometer un error de Concatenación, siendo la dificultad, una variable que influye en la probabilidad de cometer determinado tipo de error.

Siendo así, a diferencia de lo que plantea Tall (2014) y Socas(2000) por separado, ambas variables deberían ser consideradas al momento de plantear un problema dentro del contexto de una medición, si es que el objetivo es servir de insumo para la mejora de aprendizajes a través de la superación del error[1] (González, 2015).

6.3. Tipos de Error en el Contexto Pedagógico

En el caso de esta investigación, cada uno de los tipos de errores encontrados se adaptaron al contenido del álgebra lineal, de donde fueron seleccionadas las tres categorías utilizadas para el instrumento de medición. Una parte importante de esta selección fue la transformación de los tipos de errores como variable nominal a cardinal, asignándoles un valor (0, 1 y 2) que permitiera diferenciar de manera cuantitativa la magnitud o complejidad de un error respecto de otro.

Sin embargo, también es posible realizar una adaptación de categorías de errores basadas en otras tipologías que se adapten a otros contenidos o niveles, de manera que este tipo de análisis pueda ser utilizado de manera flexible tanto para mediciones estandarizadas como para retroalimentación en sala de clase o en evaluaciones a distancia.

Esto último brindaría beneficios tanto para el estudiante como para el Docente, ya que al considerar el nivel de dificultad de los ítem y el tipo de error de las

[1] Otros objetivos incluyen evaluar el estado cognitivo de los escolares o producir información útil en otros aspectos de la planificación

alternativas, el estudiante tendría la posibilidad de equivocarse conservando el nivel de logro alcanzado hasta el momento, haciendo que la evaluación del estado con fines diagnósticos (in situ o a distancia) pueda ser más "flexible" y de utilidad para su aprendizaje.

Por otro lado, al Docente le permitiría incluir sólo los ítem que se acercaran al nivel de habilidad alcanzado por los estudiantes, evaluando mediante una revisión tipo rúbrica pero automatizada aquellos aspectos específicos que necesitara reforzar con los estudiantes.

Capítulo 7

Conclusiones

En la presente investigación se estudió una tipología de errores que se adaptó para ser utilizada en la evaluación de un contenido matemático mediante un modelo de Crédito Parcial, de modo que pudiera ser aplicado como una alternativa a la evaluación clásica mediante test dicotómicos. Para ello se realizó un análisis de los tipos de error más presentes en el álgebra lineal, de los cuales se seleccionaron tres categorías: los de Concatenación, los de Yuxtaposición y los No Epistemológicos. Dadas las características de cada uno, se determinó que los de Concatenación eran errores más complejos o "profundos", y los No Epistemológicos correspondían a errores más simples o "accidentales".

Esto permitió asignarle un valor cuantitativo al error; y de igual manera en que si se hubiera seleccionado otra tipología, contenido o nivel de aprendizaje, este análisis reveló que existe una variación considerable en los tipos de errores que cometen los estudiantes frente a situaciones de desempeño con alternativas. Como por ejemplo en este caso, el error mas común fue el de Yuxtaposición con un promedio del 24 %, seguido por el de Concatenación con un 20 % y el No Epistemológico cercano al 20 % (Anexo F), brindando información útil en términos pedagógicos ya que permite comprender de mejor manera en qué se están equivocando los estudiantes.

Por otro lado, el Modelo de Crédito Parcial permitió observar que los tipos de errores estaban asociados a la dificultad del ítem, de tal manera que entre mayor fuese su dificultad existe una mayor probabilidad de cometer un error de concatenación. Por el contrario, entre menor fuese la dificultad del ítem, los datos observados indicaron que es más probable cometer los errores de yuxtaposición o No epistemológicos.

Finalmente, dado el aporte en la interpretación de los resultados de una medición estandarizada, esta metodología de construcción de instrumentos de evaluación ser utilizada con foco en la retroalimentación que se le puede otorgar a varios estudiantes de forma simultánea, como parte del apoyo técnico que necesita el Docente para poder realizarlo. De momento, sería más factible para Preuniversitarios con distintos

estudiantes en distintas sedes, pero menos factible para colegios, aunque de la podría utilizar en otros ámbitos de la matemática, y con una clase diferente de errores adecuados al contenido y nivel correspondientes.

7.1. Proyecciones

Los criterios de la validez para evaluaciones a gran escala y a nivel aula poseen límites claramente diferenciables entre sí, y en su mayoría, opuestos (Forster, C.; 2017), además, no existe concenso en la comunidad científca respecto de si es debido hablar de validez en el contexto de sala de clases (ídem), sin embargo, a partir de esta investigación podría ser posible elaborar un instrumento que pueda ser aplicado de manera masiva, bajo estándares psicométricos, pero a su vez, manteniendo una naturaleza cualitativa enfocada en apoyar al estudiante y al docente, a modo de evaluación formativa. En ese caso el siguiente paso sería elaborar una forma para retribuir la información de manera simple y práctica.

En una investigación posterior, sería relevante analizar los distintos tipos de error que se cometen en ítems que posean el mismo nivel de dificultad, verificando si ambas variables están relacionadas entre sí (dificultad y tipo de error). Esto debido a que el número de personas que respondieron este test, es relativamente bajo como para tener certeza de dicha relación.

Además, debería ser evaluado el dejar la alternativa "otro" abierta para cualquier resultado que haya obtenido el estudiante que no coincida con las alternativas, dejando abierta la posibilidad de recopilar otros tipos de errores que no estén siendo considerados por la tipología escogida. De esta manera, quedaría la opción de trabajar con 5 alternativas con sus respectivos puntajes: "otro" (0 puntos), error tipo 1 (1 punto), error tipo 2 (2 puntos), error tipo 3 (3 puntos) y respuesta correcta (4 puntos), los que con las posibilidades tecnológicas actuales pueden automatizarse para obtener los datos sin necesidad de asignar a mano cada uno de los puntajes a las respuestas. En este último caso, hay que destacar que, para que el análisis tenga sentido, los puntajes deberían asociarse a los tipos de error como en una variable de tipo ordinal, es decir, que el tipo de error más "*complejo, profundo o azaroso*" tenga asignado el puntaje 1, y el error más "*sencillo, superficial o accidental*" tenga el puntaje 3, mostrando a su vez que el estudiante tiene un mejor desempeño en esa situación problemática.

Capítulo 8

Bibliografía

Agencia de la Calidad de la Educación. (2018). *Nuevo Sistema Nacional de Evaluación de Aprendizajes.* Ministerio de Educación.

Agencia para la Cooperación y Desarrollo Económicos. (2013). *Synergies for Better Learning*. OECD Reviews of Evaluation and Assessment in Education.

Alonso, F., Barbero, C., Fuentes, I., Azcarate, A., Dozagarat, J., Gutiérrez, S., (1993). *Ideas y actividades para enseñar el álgebra.* Editorial Síntesis. Madrid. P. 50.

Anderson, L., y Krathwohl, D. (2005). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Educational HORIZONS, 83(3), 68. https://doi.org/10.1207/s15430421tip41042

Andrich, David y Marais, Ida. (2019). *A course in Rasch Measurement Theory*. Springer.

Brookhart, Susan. (2003). *Developing measurement theory for classroom assessment purposes and uses. Educational Measurement: Issues and Practice* 22:5 (12). doi: 10.1111/j.1745-3992.2003.tb00139.x.

Brookhart, Susan y Chen, Fei. (2014). *The quality and efectiveness of descriptive rubrics.* Educational Review, 67:1-26. doi:10.1080/00131911.2014.929565.

Caiceo Escudero, J. (2016). *Los sistemas estandarizados de evaluación en Chile: participación de Mario Leyton Soto y Erika Himmel König*. Historia de La Educación, 34(0), 357. https://doi.org/10.14201/hedu201534357371

Castro, E., Castro, En., Corlat, M., Marín, A., Puig, L., Sierra, M., Socas, M. (2000). *La Educación matemática en la enseñanza secundaria*. I.C.E. Universitat Barcelona. P. 126-130.

Cerda Etchepare, G., Pérez Wilson, C., Moreno Araya, C., Núñez Risco, K., Quezada Herrera, E., Rebolledo Rojas, J., y Sáez Tisnao, S. (2012). *Adaptación de la versión española del Test de Evaluación Matemática Temprana de Utrecht en Chile*. Estudios Pedagógicos (Valdivia), 38(1), 235–253. https://doi.org/10.4067/S0718-07052012000100014

D'Amore, B., Fandiño, M., Marazzani, S. y Sbaragli, S. (2014) *La didáctica y la dificultad en Matemática*. Colección Didáctica de las Matemáticas. México.

Departamento de Evaluacion medición y registro Educacional. (2017). *Evidencia preliminar para la reducción de contenidos en la PSU de matemática*.

Escudero, J. (2015). *Los sistemas estandarizados de evaluación en chile: Participación de Mario Leyton Soto y Erika Himmel Konig*. Ediciones Universidad de Salamanca, 34:357-371.

Farías, M., y Carrasco, R. (2012). *Diferencias en resultados académicos entre educación Técnico- Profesional y Humanista- Científica en Chile (Differences in academic results between Technical Professional and Scientific-Humanistic education in Chile)*. Calidad En La Educación, julio(36), 87–121. https://doi.org/10.4067/S0718-45652012000100003

Forster, C. y Rojas, C. (2008). *Evaluación al interior del aula: una mirada desde la validez, confiabilidad y objetividad.* Revista Pensamiento Educativo, 43:285-305.

Gempp, Rene. (2006). *El error estándar de medida y la puntuación verdadera de los test psicológicos.* ResearchGate, 24(02):117-129.

Hambleton, R., Swaminathan, H., y Rogers, J. (1991). *Fundamentals of Item Response Theory*. London: Sage Publications.

Hattie, J., y Timperley, H. (2007). *The power of feedback*. Review of Educational Research, 77(1), 81–112. https://doi.org/10.1111/j.1365-2923.2009.03542.x

Heitink, M., Kleij, V., Veldkamp, B., Schildkamp K., y Kippers, W. (2016). *Una revision sistemática de prerrequisitos para implementar la evaluación para el aprendizaje en la práctica de aula*. Educational Research Review 17. 50-62.

Henríquez, V. V., Gómez, C. G., Araneda, R. M., Mandiola, E. A., y Oñate, A. S. (2015). *Aprender del error es un acierto. Las dificultades que enfrentan los estudiantes chilenos en la Prueba PISA*. 293–310.

Hernández, C. A. (2017). *Diseño, desarrollo y validación para su uso en investigación, de una prueba piloto de habilidades cuantitativas para estudiantes secundarios*. Pontificia Universidad Católica de Chile.

Ibarra-González, K. P., y Eccius-Wellmann, C. (2018). *Desarrollo y Validación de un Instrumento de Medición de la Afectividad respecto a la Comisión de Errores en Matemáticas*. Bolema: Boletim de Educação Matemática, 32(61), 673–695. https://doi.org/10.1590/1980-4415v32n61a18

Kilpatrick, J., Rico, L., Sierra, M. (2011). *Educación Matemática e Investigación*. España: Editorial Síntesis.

Kilpatrick, J. (2011). *Historia de la Investigación en Educación Matemática*. España: Editorial Síntesis. pp. 17-78.

Linden, W. y Hambleton, R. (1997). *Handbook of Modern Item Response Theory*. Springer Science and Business Media.

Mannel, S. (2010). *Assessing Scientific Inquiry*. Berlin, Alemania: Logos Verlag Berlin.

Martínez Arias, R. (1996). *Psicometría: Teoría de los Test Psicológicos y Educativos*. Síntesis.

Ministerio de Educación de Chile. (2015). *Bases Curriculares 7 basico a 2 medio*. https://doi.org/10.3233/THC-171015

Morera, C. (2018). *Diseño y construcción de un instrumento de evaluación de la competencia matemática: aplicabilidad práctica de un juicio de expertos*. 347–372.

Morgado, Alonso. (2016). *El preuniversitario: un ejemplo de educación no formal. El caso del preuniversitario Pedro de Valdivia*. Revista Enfoques Educacionales, 13(1):91-102.

National Council of Teachers of Mathematics NCTM. (2011). *Expressions, equations and Functions*. Reston, Va.

Nicol, D. y Macfarlane, D. (2006). *Formative assessment and self-regulated learning: A model and seven principles of good feedback practice*. Studies in Higher Education, 31(2):199-218.

Ortiz, A. (2015). *Enseñanza, Aprendizaje y Evaluación: una configuración triádica*. Proyecto de investigación MODEPED. Revista TED N°37, pp. 87 – 108.

Pedag, E., Naturales, C., y Valpara, P. A. (2010). *Didactics on Knowledge: a Road to the Optimization of Didactic Transpositions*. (1991), 51–61.

Preiss, D. D., Calcagni, E., Espinoza, A. M., Gómez, D., Grau, V., Guzmán, V., Ramírez, F. (2014). *Buenas Prácticas Pedagógicas Observadas en el Aula de Segundo Ciclo Básico en Chile*. 23, 1–12. https://doi.org/10.7764/psykhe.23.2.716

Radatz, H. (1979). *Error Analysis in Mathematics Education*. National Council of Teachers of Mathematics. Journal for Research in Mathematics Education, Vol. 10, No. 3.

Rico, L. (2000). *La Educación Matemática en la Enseñanza Secundaria*. Barcelona: Universitat Barcelona.

Robins, R., Fraley, C., y Krueger, R. (2007). *Handbook of Research Methods in personality Phychology*. London: The Guilford Press.

Rosso, A., Barros, J. (2013). *Una Taxonomía de errores en el aprendizaje de espacio vectoriales*. Argentina: Revista Iberoamericanas para la Educación, la Ciencia y la Tecnología. ISSN: 1681-5653 n° 63/2

Ruiz, G. (2013). *La teoría de la experiencia de John Dewey: significación histórica y vigencia en el debate teórico contemporáneo*. Foro de Educación, 11(15), 103–124. https://doi.org/10.14516/fde.2013.011.015.005

Sadler, D.R. (1989). *Formative assessment and the desing of instructional systems.* Instructional Science, 18:119-144.

Socas, M. (2000). *Dificultades, obstáculos y errores en el aprendizaje de las Matemáticas en la Educación Secundaria*. Barcelona. Universitat of Barcelona. pp. 125-154.

SUMMA. (s.n.). *Retroalimentación formativa*. Disponible en https://www.summaedu.org /retroalimentacion-formativa/

SUMMA. (2018). *Jornada de Reflexión sobre prácticas educativas efectivas*. Santiago. Conferencia Orientaciones para la Formación Docente y el Trabajo en el Aula.

Tall, D. (2014). *New Cognitive Obstacles in a Technological Paradigm*. Mathematics Education Research Center, University of Warwick. Obtenido de https://homepages. warwick.ac.uk/staff/David.Tall/pdfs/dot1989b-cog-obst-nctm.pdf

Taut, S., Manzi, J., y Molina, A. (2013). *La evaluación de las prácticas docentes. Pensamiento Educativo*. Revista de Investigación Educacional Latinoamericana, 50(1), 1–3. https://doi.org/10.7764/pel.50.1.2013.1

Treviño Villarreal, E. (2006). *Evaluación del aprendizaje de los estudiantes indígenas en América Latina*. Revista Mexicana de Investigación Educativa, 11(28), 225–268.

Treviño, E., Villalobos, C. y Baeza, A. (2016). *Recomendaciones de Políticas Educativas en América Latina en base al TERCE*. Organización de las Naciones Unidas para la Educación la Ciencia y la Cultura.

UNESCO Oficina Internacional de Educación (2004). *La Educación Chilena en el cambio de Siglo: Políticas, Resultados y Desafíos*. Ministerio de Educación.

Villamil, L. E. (2008). *Revista de Estudios Literarios*. (U. C. Madrid, Ed.) Obtenido de Espéculo N° 38: https://webs.ucm.es/info/especulo/numero38/obstepis.html

William, D. (2009). *Una síntesis integradora de la investigación e implicancias para una nueva teoría de la evaluación formativa*. Archivo de Ciencias de la Educación (4ª. Época), año 3, No. 3, p15-44.

Wu, Margaret., Ping Tam, Hak y Jen, Tsung-Hau. (2016). *Educational Measurement for Applied Researchers*. Singapore: Springer.

Anexo A

Parámetros estimados del ítem

A.1. Dificultad por ítem

Figura A.1: Fuente: Datos obtenidos con Conquest

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item

	VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
	item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	N1	-0.191	0.082	0.98	(0.73, 1.27)	-0.1	0.95	(0.82, 1.18)	-0.6
2	N2	0.131	0.063	0.75	(0.73, 1.27)	-1.9	0.80	(0.84, 1.16)	-2.7
3	N3	0.364	0.068	0.95	(0.73, 1.27)	-0.3	0.96	(0.80, 1.20)	-0.4
4	N4	-0.218	0.068	0.75	(0.73, 1.27)	-1.9	0.85	(0.78, 1.22)	-1.4
5	N5	-0.068	0.065	1.32	(0.73, 1.27)	2.2	1.24	(0.83, 1.17)	2.6
6	N6	0.024	0.065	1.13	(0.73, 1.27)	0.9	1.10	(0.84, 1.16)	1.2
7	N7	-0.030	0.074	1.01	(0.73, 1.27)	0.1	0.99	(0.80, 1.20)	-0.0
8	N8	-0.404	0.071	1.28	(0.73, 1.27)	1.9	1.14	(0.82, 1.18)	1.5
9	N9	-0.445	0.070	0.83	(0.73, 1.27)	-1.3	0.89	(0.82, 1.18)	-1.2
10	N10	0.281	0.067	1.13	(0.73, 1.27)	0.9	1.12	(0.82, 1.18)	1.3
11	N11	0.171	0.068	1.02	(0.73, 1.27)	0.2	1.02	(0.82, 1.18)	0.3
12	N12	0.027	0.071	1.07	(0.73, 1.27)	0.6	1.07	(0.81, 1.19)	0.7
13	N13	0.389	0.068	1.21	(0.73, 1.27)	1.5	1.20	(0.83, 1.17)	2.2
14	N14	-0.129	0.070	0.91	(0.73, 1.27)	-0.6	0.92	(0.82, 1.18)	-0.8
15	N15	-0.172	0.068	0.78	(0.73, 1.27)	-1.7	0.86	(0.80, 1.20)	-1.5
16	N16	0.007	0.075	0.93	(0.73, 1.27)	-0.5	0.92	(0.79, 1.21)	-0.7
17	N17	-0.137	0.072	1.09	(0.73, 1.27)	0.7	1.08	(0.82, 1.18)	0.8
18	N18	0.204	0.074	1.00	(0.73, 1.27)	0.0	0.99	(0.79, 1.21)	-0.0
19	N19	0.638	0.068	1.00	(0.73, 1.27)	0.0	0.94	(0.80, 1.20)	-0.6
20	N20	-0.570	0.074	0.88	(0.73, 1.27)	-0.9	0.92	(0.79, 1.21)	-0.7
21	N21	0.253	0.068	1.07	(0.73, 1.27)	0.5	1.04	(0.82, 1.18)	0.4
22	N22	0.604	0.067	1.68	(0.73, 1.27)	4.1	1.59	(0.80, 1.20)	5.1
23	N23	-0.024	0.064	0.96	(0.73, 1.27)	-0.3	0.97	(0.84, 1.16)	-0.4
24	N24	-0.287	0.074	0.88	(0.73, 1.27)	-0.9	0.89	(0.79, 1.21)	-1.0
25	N25	-0.479	0.074	0.83	(0.73, 1.27)	-1.3	0.91	(0.73, 1.27)	-0.6
26	N26	-0.252	0.074	0.83	(0.73, 1.27)	-1.2	0.85	(0.80, 1.20)	-1.5
27	N27	0.370	0.065	0.73	(0.73, 1.27)	-2.1	0.76	(0.81, 1.19)	-2.7
28	N28	-0.299	0.072	0.84	(0.73, 1.27)	-1.2	0.86	(0.83, 1.17)	-1.7
29	N29	0.370	0.081	1.18	(0.73, 1.27)	1.3	1.16	(0.74, 1.26)	1.2
30	N30	-0.129*	0.380	0.96	(0.73, 1.27)	-0.2	0.98	(0.77, 1.23)	-0.2

An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.952
Chi-square test of parameter equality = 582.74, df = 29, Sig Level = 0.000
^ Quick standard errors have been used

A.2. Dificultad según tipo de error

Figura A.2: Fuente: Datos obtenidos con Conquest

TERM 2: item*step

VARIABLES					UNWEIGHTED FIT			WEIGHTED FIT		
	item	step	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	N1	1			1.14	(0.73, 1.27)	1.0	1.04	(0.73, 1.27)	0.3
1	N1	2	-0.179	0.206	0.93	(0.73, 1.27)	-0.5	0.95	(0.87, 1.13)	-0.8
1	N1	3	0.179*		0.88	(0.73, 1.27)	-0.8	0.89	(0.91, 1.09)	-2.5
2	N2	0			0.84	(0.73, 1.27)	-1.2	0.93	(0.85, 1.15)	-0.9
2	N2	1	0.878	0.245	0.86	(0.73, 1.27)	-1.0	0.96	(0.62, 1.38)	-0.1
2	N2	2	1.003	0.421	0.98	(0.73, 1.27)	-0.1	1.00	(0.27, 1.73)	0.1
2	N2	3	-1.881*		0.74	(0.73, 1.27)	-2.1	0.77	(0.86, 1.14)	-3.5
3	N3	0			1.09	(0.73, 1.27)	0.7	1.09	(0.86, 1.14)	1.2
3	N3	1	-0.019	0.200	0.93	(0.73, 1.27)	-0.5	0.96	(0.82, 1.18)	-0.5
3	N3	2	0.811	0.307	1.13	(0.73, 1.27)	0.9	1.03	(0.54, 1.46)	0.2
3	N3	3	-0.792*		0.84	(0.73, 1.27)	-1.2	0.84	(0.76, 1.24)	-1.4
4	N4	0			0.73	(0.73, 1.27)	-2.2	0.90	(0.69, 1.31)	-0.6
4	N4	1	1.009	0.225	0.87	(0.73, 1.27)	-0.9	0.98	(0.45, 1.55)	0.0
4	N4	2	-0.301	0.259	0.95	(0.73, 1.27)	-0.4	0.98	(0.67, 1.33)	-0.1
4	N4	3	-0.708*		0.80	(0.73, 1.27)	-1.5	0.84	(0.89, 1.11)	-3.0
5	N5	0			1.56	(0.73, 1.27)	3.5	1.24	(0.77, 1.23)	1.9
5	N5	1	0.652	0.225	0.93	(0.73, 1.27)	-0.5	0.98	(0.62, 1.38)	-0.1
5	N5	2	0.484	0.307	0.98	(0.73, 1.27)	-0.1	0.99	(0.54, 1.46)	0.1
5	N5	3	-1.136*		1.10	(0.73, 1.27)	0.8	1.12	(0.88, 1.12)	1.8
6	N6	0			1.00	(0.73, 1.27)	0.0	1.03	(0.80, 1.20)	0.3
6	N6	1	0.506	0.217	1.24	(0.73, 1.27)	1.6	1.04	(0.68, 1.32)	0.3
6	N6	2	0.588	0.307	1.02	(0.73, 1.27)	0.2	1.01	(0.54, 1.46)	0.1
6	N6	3	-1.094*		1.06	(0.73, 1.27)	0.5	1.05	(0.87, 1.13)	0.7
7	N7	0			0.86	(0.73, 1.27)	-1.0	0.98	(0.56, 1.44)	0.0
7	N7	1	-1.152	0.205	1.00	(0.73, 1.27)	0.1	1.01	(0.92, 1.08)	0.2
7	N7	2	1.022	0.244	1.03	(0.73, 1.27)	0.2	1.01	(0.71, 1.29)	0.1
7	N7	3	0.130*		1.05	(0.73, 1.27)	0.4	0.98	(0.77, 1.23)	-0.2
8	N8	0			2.42	(0.73, 1.27)	7.5	1.08	(0.47, 1.53)	0.4
8	N8	1	-0.694	0.204	1.02	(0.73, 1.27)	0.2	1.02	(0.82, 1.18)	0.2
8	N8	2	1.504	0.307	0.98	(0.73, 1.27)	-0.1	0.99	(0.54, 1.46)	0.1
8	N8	3	-0.810*		1.05	(0.73, 1.27)	0.4	1.05	(0.89, 1.11)	0.9
9	N9	0			0.89	(0.73, 1.27)	-0.8	1.00	(0.47, 1.53)	0.1
9	N9	1	-0.657	0.211	0.83	(0.73, 1.27)	-1.2	0.91	(0.82, 1.18)	-1.0
9	N9	2	2.232	0.421	0.95	(0.73, 1.27)	-0.3	0.99	(0.27, 1.73)	0.1
9	N9	3	-1.575*		0.82	(0.73, 1.27)	-1.4	0.86	(0.88, 1.12)	-2.5
10	N10	1	0.319	0.200	1.02	(0.73, 1.27)	0.2	1.01	(0.72, 1.28)	0.1
10	N10	2	-0.042	0.245	1.05	(0.73, 1.27)	0.4	1.03	(0.71, 1.29)	0.2
10	N10	3	-0.278*		1.07	(0.73, 1.27)	0.6	1.03	(0.79, 1.21)	0.3
11	N11	0			1.07	(0.73, 1.27)	0.6	1.08	(0.80, 1.20)	0.8
11	N11	1	0.086	0.198	1.00	(0.73, 1.27)	0.1	0.99	(0.76, 1.24)	-0.1
11	N11	2	0.190	0.244	1.05	(0.73, 1.27)	0.4	1.03	(0.71, 1.29)	0.2
11	N11	3	-0.277*		0.90	(0.73, 1.27)	-0.7	0.90	(0.81, 1.19)	-1.0
12	N12	0			1.12	(0.73, 1.27)	0.9	1.02	(0.69, 1.31)	0.2
12	N12	1	-0.406	0.197	0.99	(0.73, 1.27)	-0.0	0.97	(0.82, 1.18)	-0.3
12	N12	2	0.415	0.233	0.96	(0.73, 1.27)	-0.2	0.98	(0.75, 1.25)	-0.1
12	N12	3	-0.009*		1.14	(0.73, 1.27)	1.0	1.13	(0.82, 1.18)	1.4
13	N13	0			1.08	(0.73, 1.27)	0.6	1.09	(0.86, 1.14)	1.2
13	N13	1	0.861	0.198	1.14	(0.73, 1.27)	1.0	1.02	(0.54, 1.46)	0.2
13	N13	2	-1.157	0.211	1.07	(0.73, 1.27)	0.6	1.04	(0.84, 1.16)	0.5
13	N13	3	0.296*		1.26	(0.73, 1.27)	1.8	1.15	(0.74, 1.26)	1.1
14	N14	0			0.77	(0.73, 1.27)	-1.8	0.93	(0.64, 1.36)	-0.3
14	N14	1	-0.443	0.198	1.01	(0.73, 1.27)	0.1	1.00	(0.82, 1.18)	0.1
14	N14	2	0.844	0.259	1.01	(0.73, 1.27)	0.1	1.00	(0.67, 1.33)	0.1
14	N14	3	-0.401*		0.92	(0.73, 1.27)	-0.5	0.92	(0.86, 1.14)	-1.1
15	N15	0			0.72	(0.73, 1.27)	-2.2	0.90	(0.69, 1.31)	-0.6
15	N15	1	0.461	0.212	0.87	(0.73, 1.27)	-0.9	0.97	(0.62, 1.38)	-0.1
15	N15	2	0.175	0.259	0.96	(0.73, 1.27)	-0.3	0.98	(0.67, 1.33)	-0.0
15	N15	3	-0.636*		0.85	(0.73, 1.27)	-1.1	0.87	(0.89, 1.11)	-2.4

Figura A.3: Fuente: Datos obtenidos con Conquest

```
16  N16             0                          0.79 ( 0.73, 1.27) -1.6   0.94 ( 0.56, 1.44) -0.2
16  N16             1            -1.049   0.207    1.02 ( 0.73, 1.27)  0.2   1.03 ( 0.89, 1.11)  0.5
16  N16             2             0.483   0.216    1.04 ( 0.73, 1.27)  0.3   1.03 ( 0.82, 1.18)  0.3
16  N16             3             0.566*           0.85 ( 0.73, 1.27) -1.1   0.87 ( 0.74, 1.26) -1.0
17  N17             0                          1.69 ( 0.73, 1.27)  4.2   1.08 ( 0.56, 1.44)  0.4
17  N17             1            -1.053   0.200    0.93 ( 0.73, 1.27) -0.5   0.95 ( 0.92, 1.08) -1.2
17  N17             2             1.678   0.307    1.03 ( 0.73, 1.27)  0.3   1.01 ( 0.54, 1.46)  0.1
17  N17             3            -0.625*           1.02 ( 0.73, 1.27)  0.2   0.97 ( 0.83, 1.17) -0.4
18  N18             0                          1.11 ( 0.73, 1.27)  0.8   1.04 ( 0.68, 1.32)  0.3
18  N18             1            -0.901   0.203    0.97 ( 0.73, 1.27) -0.2   0.98 ( 0.91, 1.09) -0.4
18  N18             2             0.601   0.233    1.11 ( 0.73, 1.27)  0.8   1.05 ( 0.75, 1.25)  0.5
18  N18             3             0.301*           0.86 ( 0.73, 1.27) -1.0   0.88 ( 0.71, 1.29) -0.8
19  N19             0                          1.00 ( 0.73, 1.27)  0.1   1.02 ( 0.88, 1.12)  0.3
19  N19             1             0.989   0.207    0.97 ( 0.73, 1.27) -0.2   0.99 ( 0.54, 1.46)  0.0
19  N19             2            -1.050   0.235    1.10 ( 0.73, 1.27)  0.7   1.07 ( 0.75, 1.25)  0.6
19  N19             3             0.062*           1.00 ( 0.73, 1.27)  0.0   0.83 ( 0.67, 1.33) -1.0
20  N20             0                          0.82 ( 0.73, 1.27) -1.3   0.96 ( 0.30, 1.70)  0.0
20  N20             1            -0.722   0.203    0.92 ( 0.73, 1.27) -0.6   0.98 ( 0.76, 1.24) -0.2
20  N20             2             1.029   0.259    0.95 ( 0.73, 1.27) -0.3   0.98 ( 0.67, 1.33) -0.1
20  N20             3            -0.307*           0.87 ( 0.73, 1.27) -0.9   0.90 ( 0.90, 1.10) -2.1
21  N21             0                          0.92 ( 0.73, 1.27) -0.5   1.00 ( 0.82, 1.18) -0.0
21  N21             1            -0.102   0.198    0.99 ( 0.73, 1.27)  0.0   1.00 ( 0.82, 1.18)  0.0
21  N21             2             0.684   0.279    1.04 ( 0.73, 1.27)  0.4   1.02 ( 0.61, 1.39)  0.2
21  N21             3            -0.581*           1.18 ( 0.73, 1.27)  1.3   1.02 ( 0.79, 1.21)  0.2
22  N22             0                          1.60 ( 0.73, 1.27)  3.7   1.47 ( 0.87, 1.13)  6.4
22  N22             1             1.367   0.217    0.94 ( 0.73, 1.27) -0.4   0.99 ( 0.44, 1.56)  0.1
22  N22             2            -1.172   0.246    1.12 ( 0.73, 1.27)  0.9   1.06 ( 0.71, 1.29)  0.5
22  N22             3            -0.196*           1.77 ( 0.73, 1.27)  4.6   1.38 ( 0.71, 1.29)  2.3
23  N23             0                          0.82 ( 0.73, 1.27) -1.3   0.95 ( 0.80, 1.20) -0.5
23  N23             1             0.729   0.234    1.16 ( 0.73, 1.27)  1.2   1.02 ( 0.62, 1.38)  0.2
23  N23             2             0.733   0.349    1.00 ( 0.73, 1.27)  0.1   1.00 ( 0.43, 1.57)  0.1
23  N23             3            -1.462*           0.95 ( 0.73, 1.27) -0.3   0.94 ( 0.88, 1.12) -0.9
24  N24             0                          0.70 ( 0.73, 1.27) -2.4   0.93 ( 0.46, 1.54) -0.2
24  N24             1            -0.585   0.199    1.00 ( 0.73, 1.27)  0.1   1.02 ( 0.76, 1.24)  0.2
24  N24             2             0.161   0.210    0.99 ( 0.73, 1.27) -0.1   0.99 ( 0.85, 1.15) -0.1
24  N24             3             0.424*           0.83 ( 0.73, 1.27) -1.2   0.84 ( 0.86, 1.14) -2.3
25  N25             0                          0.89 ( 0.73, 1.27) -0.8   1.01 ( 0.47, 1.53)  0.1
25  N25             1             0.276   0.213    0.75 ( 0.73, 1.27) -1.9   0.93 ( 0.55, 1.45) -0.2
25  N25             2             0.081   0.245    0.94 ( 0.73, 1.27) -0.4   0.97 ( 0.72, 1.28) -0.1
25  N25             3            -0.357*           0.85 ( 0.73, 1.27) -1.1   0.89 ( 0.89, 1.11) -2.1
26  N26             0                          0.81 ( 0.73, 1.27) -1.4   0.96 ( 0.46, 1.54) -0.0
26  N26             1            -0.837   0.200    0.87 ( 0.73, 1.27) -0.9   0.93 ( 0.82, 1.18) -0.7
26  N26             2             0.453   0.216    0.99 ( 0.73, 1.27)  0.0   1.00 ( 0.82, 1.18) -0.0
26  N26             3             0.384*           0.84 ( 0.73, 1.27) -1.2   0.86 ( 0.84, 1.16) -1.8
27  N27             0                          0.87 ( 0.73, 1.27) -0.9   0.93 ( 0.88, 1.12) -1.3
27  N27             1             0.555   0.224    0.92 ( 0.73, 1.27) -0.5   0.97 ( 0.72, 1.28) -0.2
27  N27             2             1.140   0.421    1.10 ( 0.73, 1.27)  0.8   1.02 ( 0.27, 1.73)  0.2
27  N27             3            -1.696*           0.66 ( 0.73, 1.27) -2.8   0.70 ( 0.79, 1.21) -3.1
28  N28             0                          0.94 ( 0.73, 1.27) -0.4   1.00 ( 0.46, 1.54)  0.1
28  N28             1            -1.121   0.200    0.88 ( 0.73, 1.27) -0.9   0.91 ( 0.90, 1.10) -1.8
28  N28             2             1.757   0.307    1.01 ( 0.73, 1.27)  0.1   1.00 ( 0.54, 1.46)  0.1
28  N28             3            -0.636*           0.80 ( 0.73, 1.27) -1.5   0.81 ( 0.86, 1.14) -2.8
29  N29             0                          0.95 ( 0.73, 1.27) -0.4   1.00 ( 0.55, 1.45)  0.1
29  N29             1            -1.713   0.244    1.07 ( 0.73, 1.27)  0.5   1.07 ( 0.95, 1.05)  2.6
29  N29             2             0.481   0.218    1.04 ( 0.73, 1.27)  0.3   1.03 ( 0.81, 1.19)  0.3
29  N29             3             1.232*           1.58 ( 0.73, 1.27)  3.6   1.16 ( 0.44, 1.56)  0.6
30  N30             0                          0.90 ( 0.73, 1.27) -0.7   0.97 ( 0.56, 1.44) -0.1
30  N30             1            -0.162   0.199    0.94 ( 0.73, 1.27) -0.4   0.98 ( 0.68, 1.32) -0.1
30  N30             2            -0.520   0.199    1.01 ( 0.73, 1.27)  0.1   1.01 ( 0.92, 1.08)  0.3
30  N30             3             0.682*           0.99 ( 0.73, 1.27) -0.0   1.01 ( 0.82, 1.18)  0.1
-------------------------------------------------------------------------------------------------------------------
            An asterisk next to a parameter estimate indicates that it is constrained
            ^ Quick standard errors have been used
===================================================================================================================
```

Anexo B

Mapas de Distribución

Figura B.1: Mapa de Dificultad de los ítems

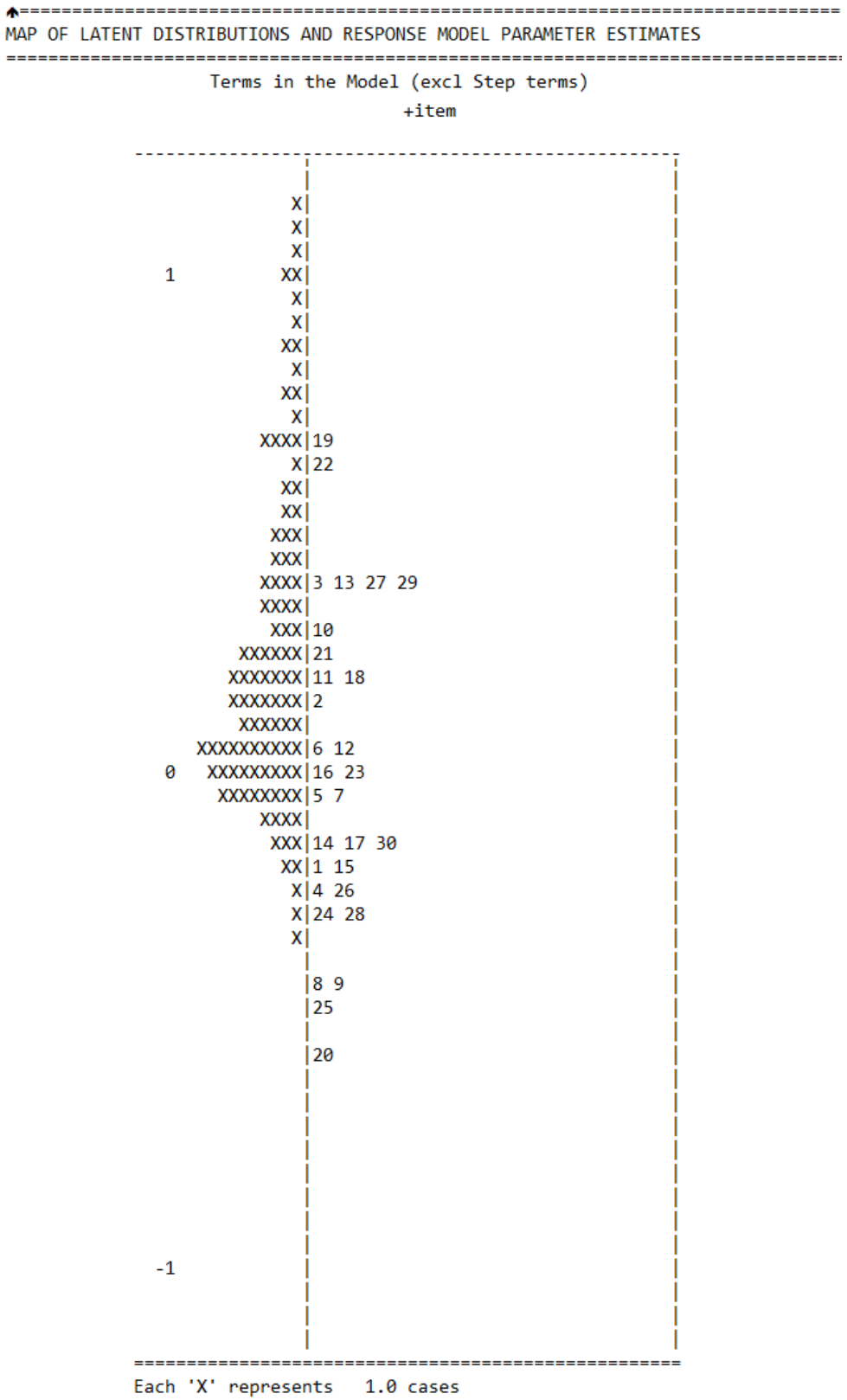


Figura B.2: Mapa de Dificultad según tipo de error

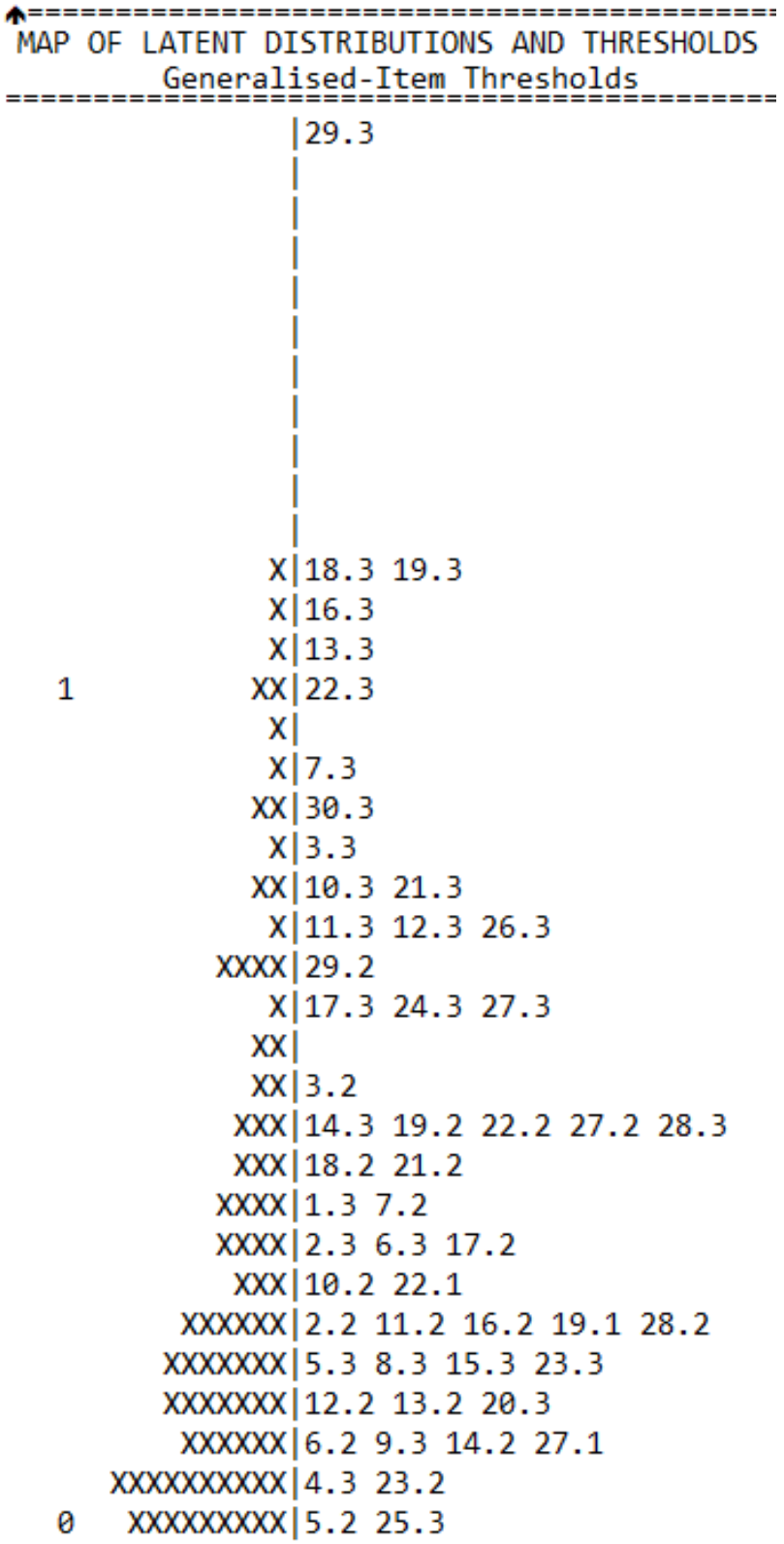
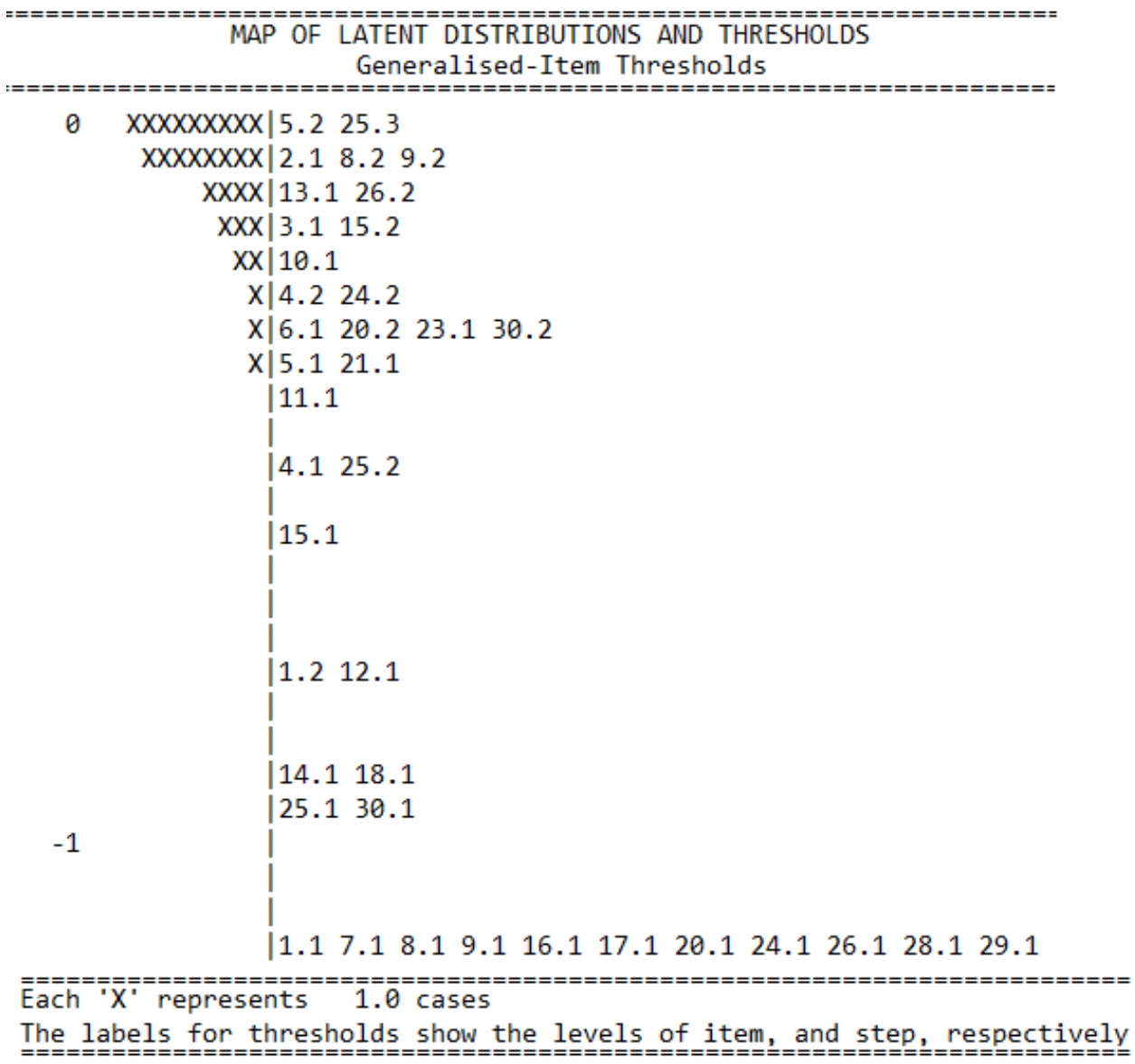


Figura B.3: Mapa de Dificultad según tipo de error



Anexo C

Curvas Características del ítem

Figura C.1: Curva Característica del ítem N°1

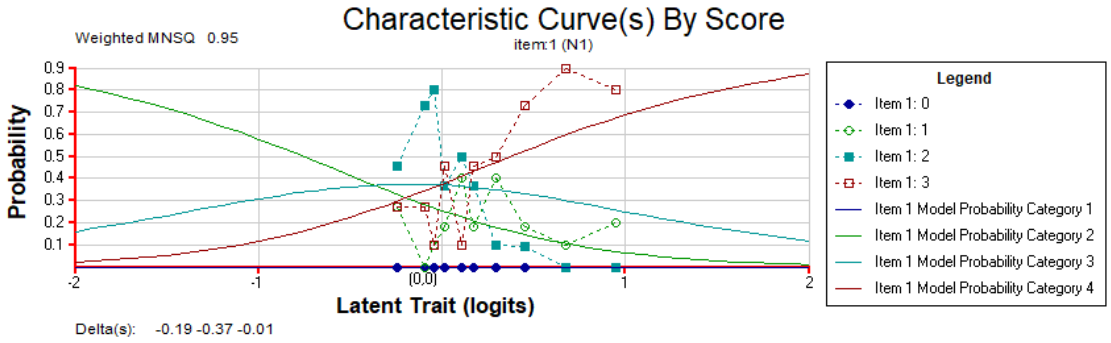


Figura C.2: Curva Característica del ítem N°2

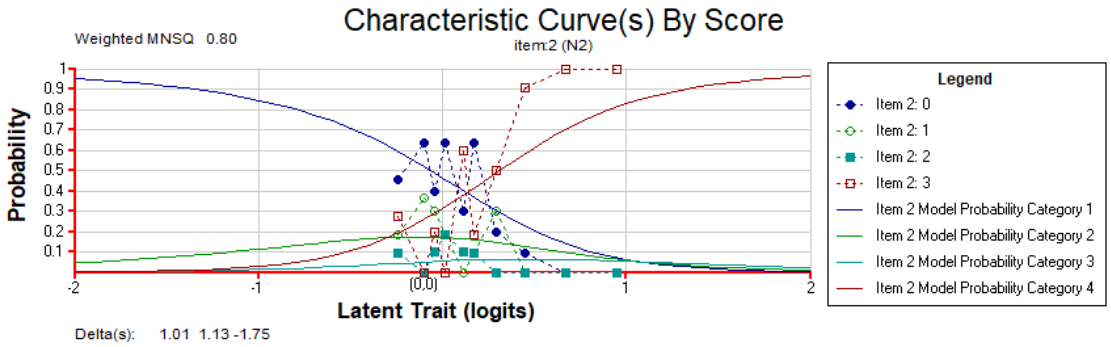


Figura C.3: Curva Característica del ítem N°3

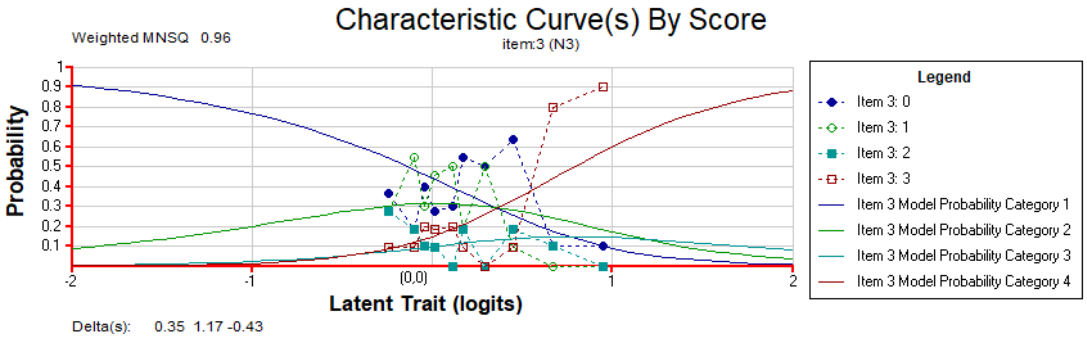


Figura C.4: Curva Característica del ítem N°4

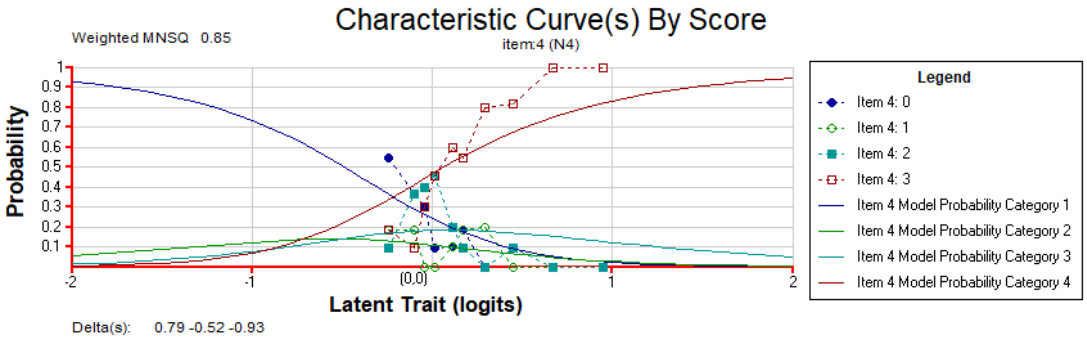


Figura C.5: Curva Característica del ítem N°5

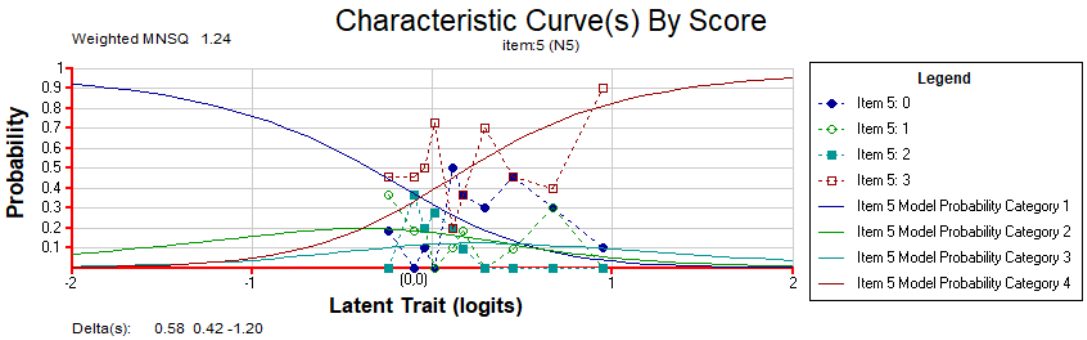


Figura C.6: Curva Característica del ítem N°6

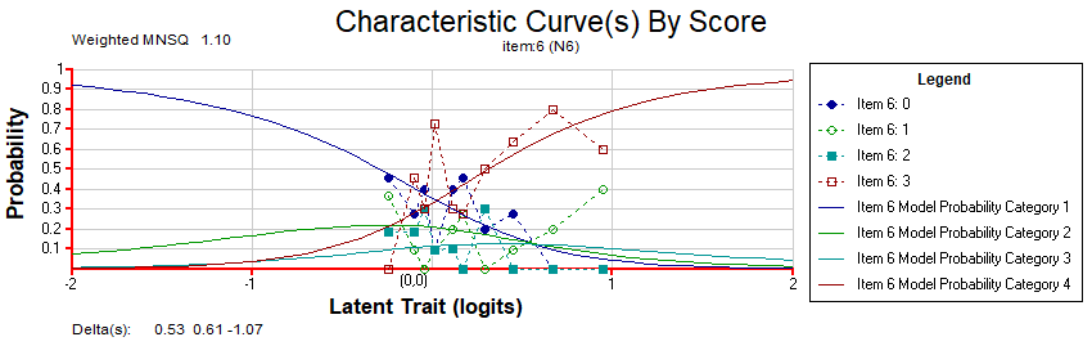


Figura C.7: Curva Característica del ítem N°7

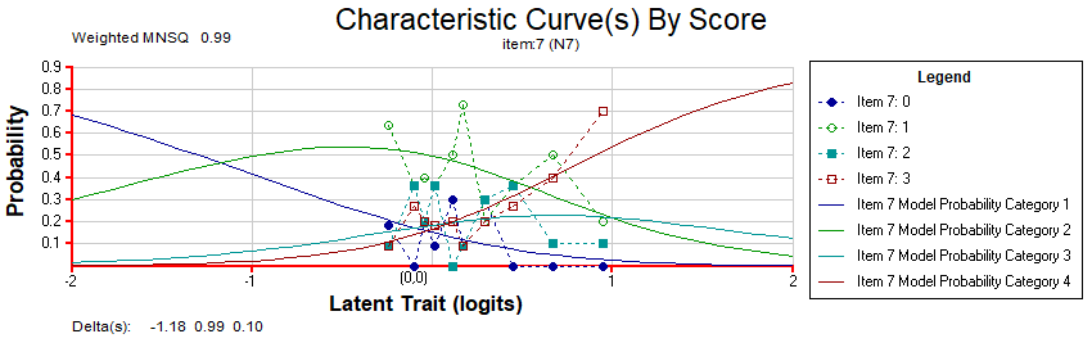


Figura C.8: Curva Característica del ítem N°8

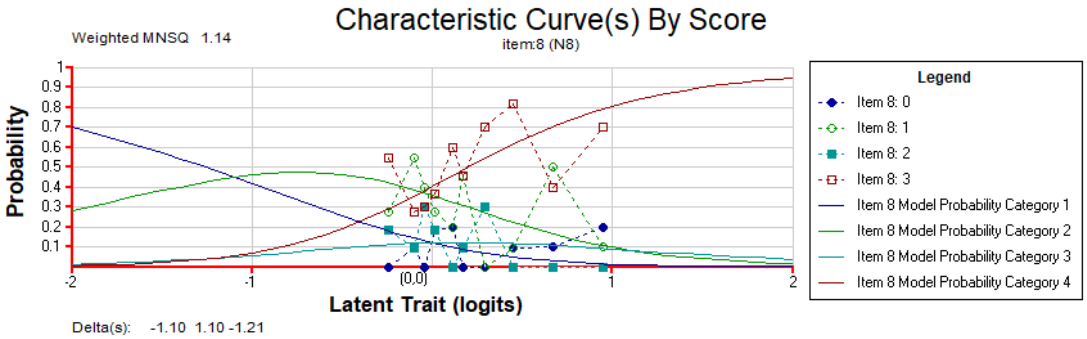


Figura C.9: Curva Característica del ítem N°9

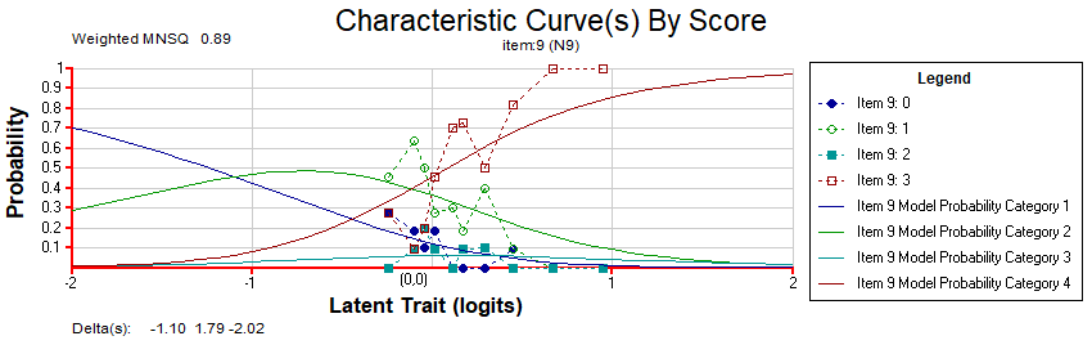


Figura C.10: Curva Característica del ítem N°10

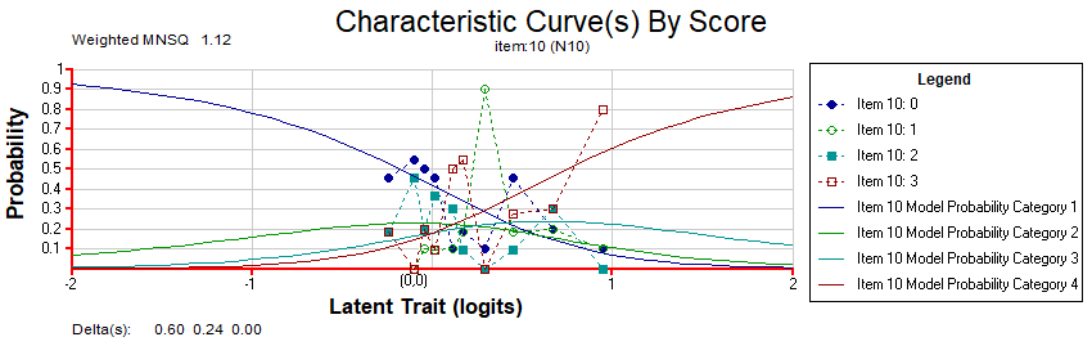


Figura C.11: Curva Característica del ítem N°11

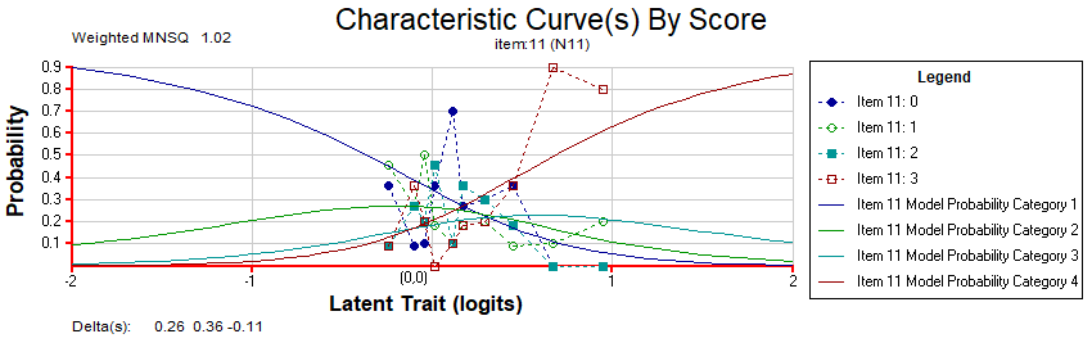


Figura C.12: Curva Característica del ítem N°12

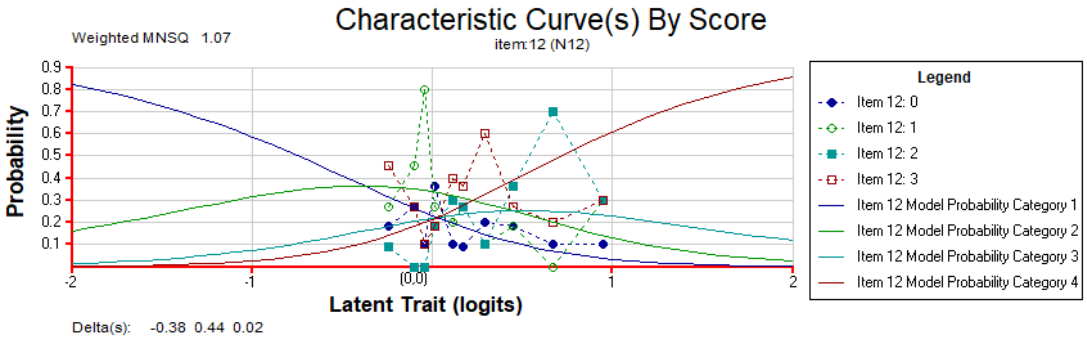


Figura C.13: Curva Característica del ítem N°13

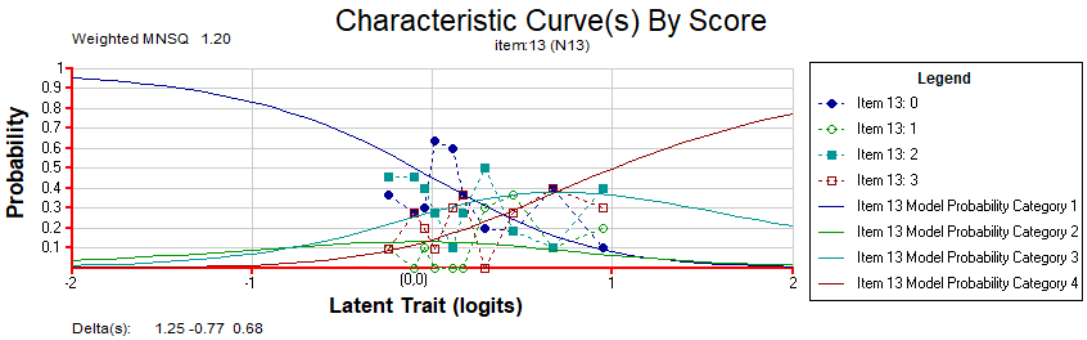


Figura C.14: Curva Característica del ítem N°14

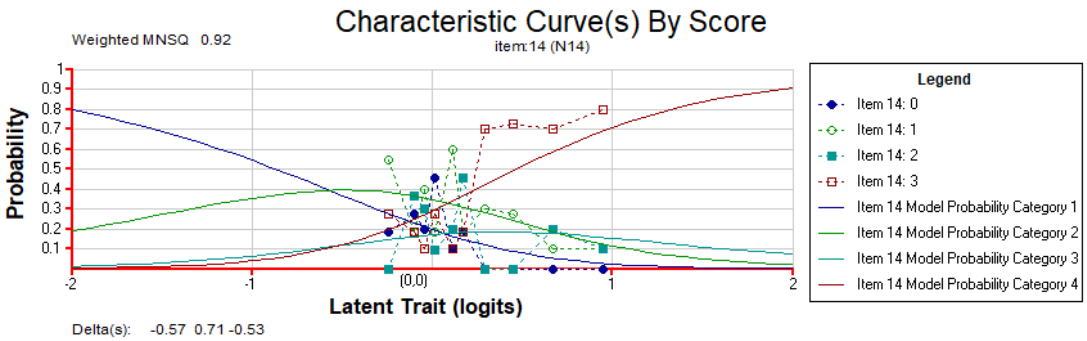


Figura C.15: Curva Característica del ítem N°15

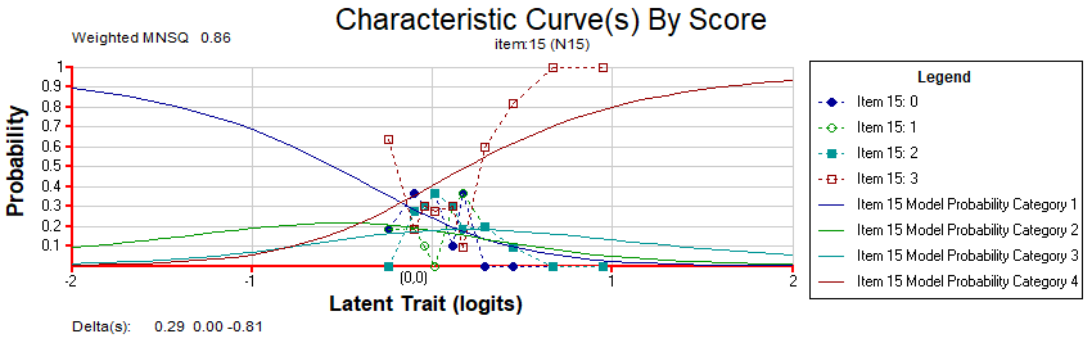


Figura C.16: Curva Característica del ítem N°16

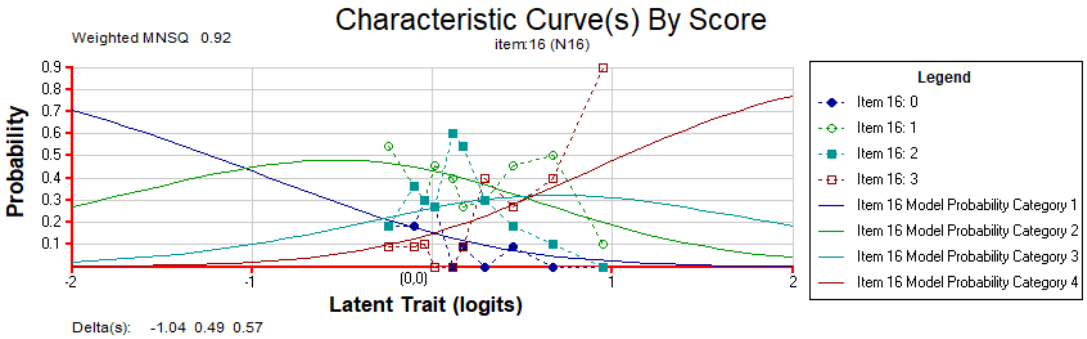


Figura C.17: Curva Característica del ítem N°17

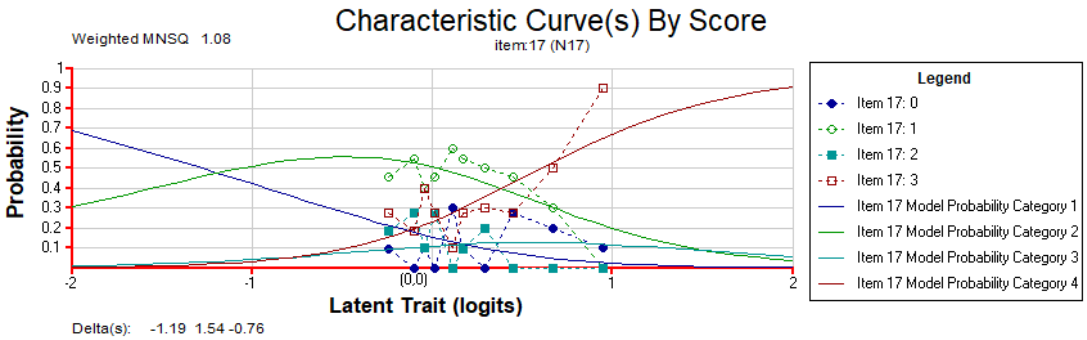


Figura C.18: Curva Característica del ítem N°18

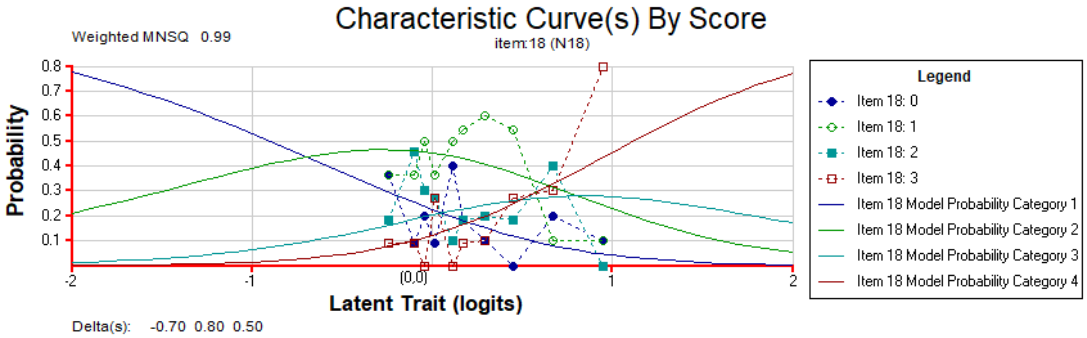


Figura C.19: Curva Característica del ítem N°19

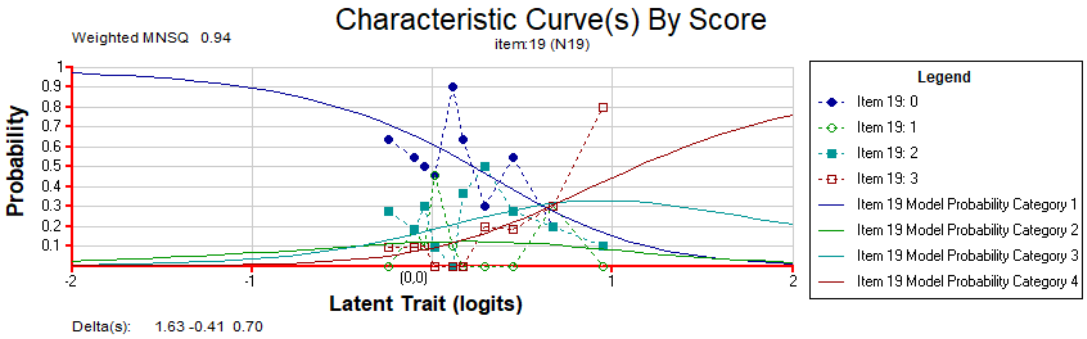


Figura C.20: Curva Característica del ítem N°20

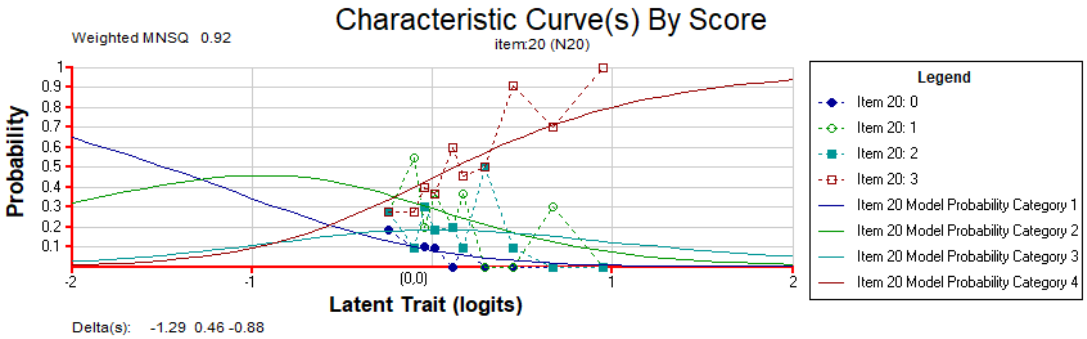


Figura C.21: Curva Característica del ítem N°21

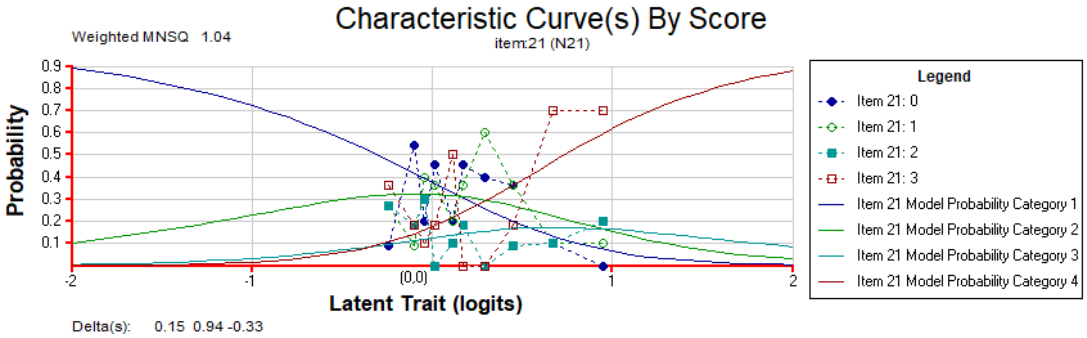


Figura C.22: Curva Característica del ítem N°22

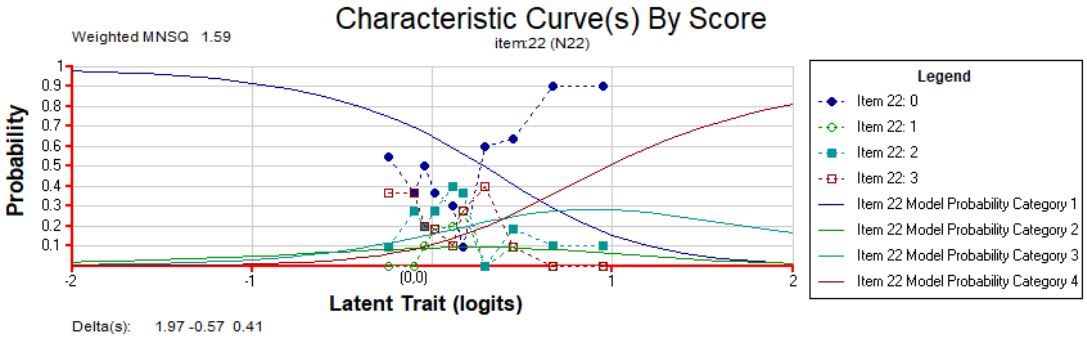


Figura C.23: Curva Característica del ítem N°23

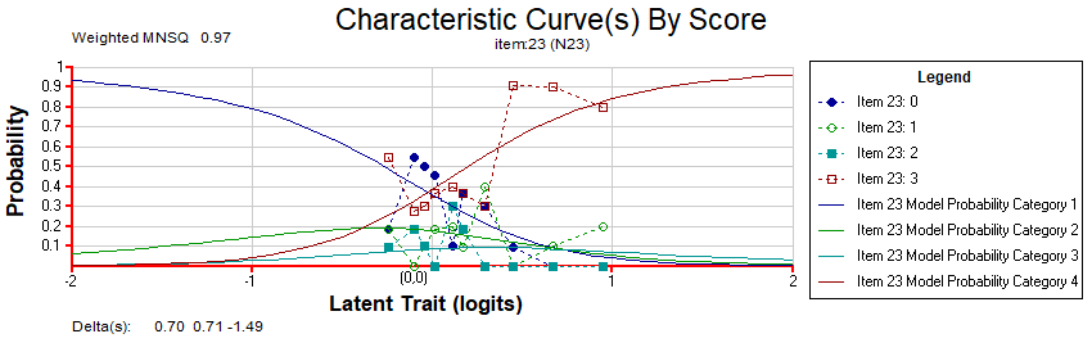


Figura C.24: Curva Característica del ítem N°24

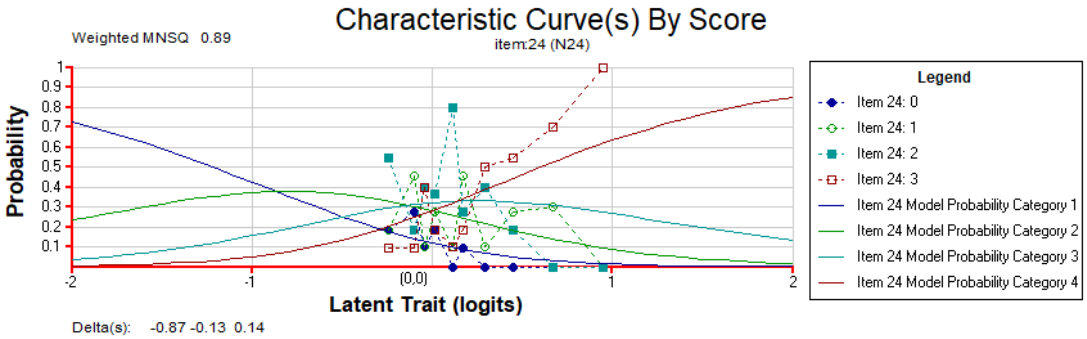


Figura C.25: Curva Característica del ítem N°25

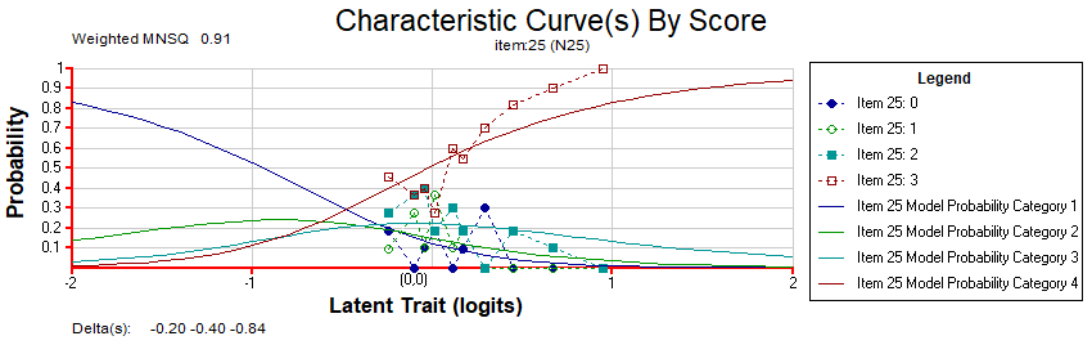


Figura C.26: Curva Característica del ítem N°26

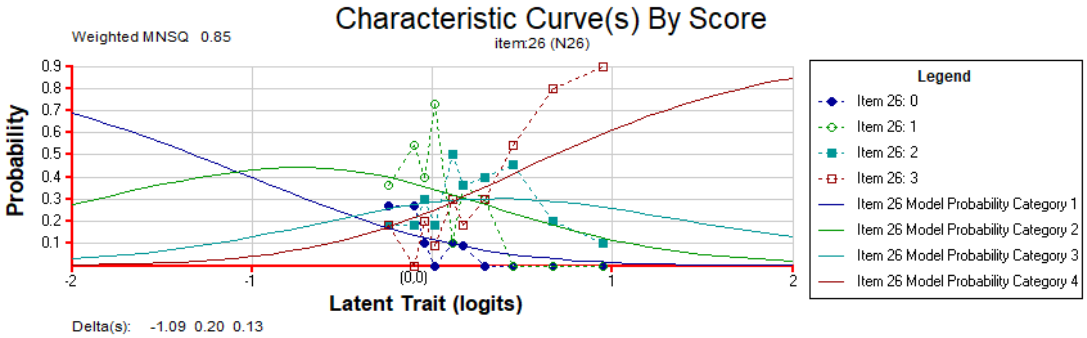


Figura C.27: Curva Característica del ítem N°27

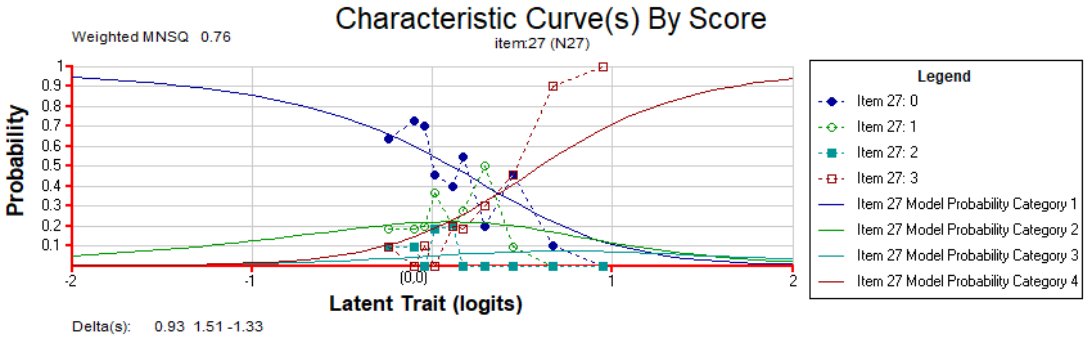


Figura C.28: Curva Característica del ítem N°28

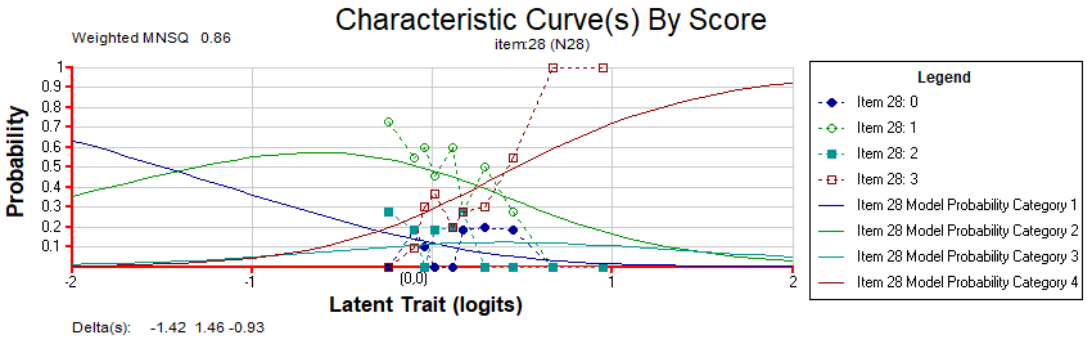


Figura C.29: Curva Característica del ítem N°29

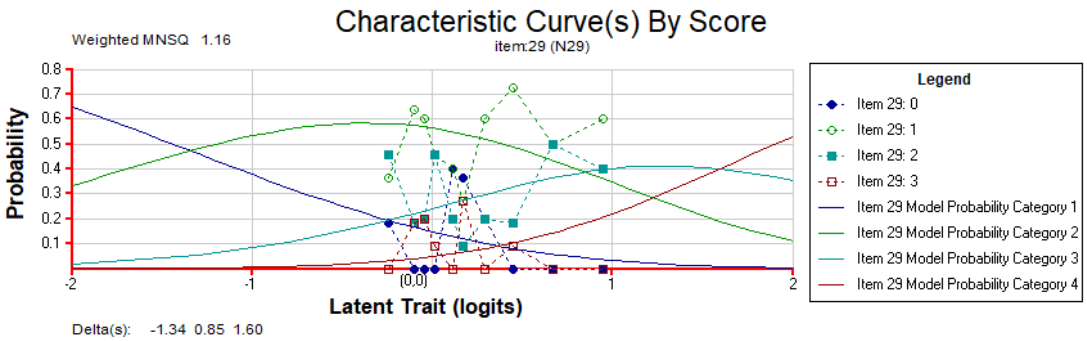
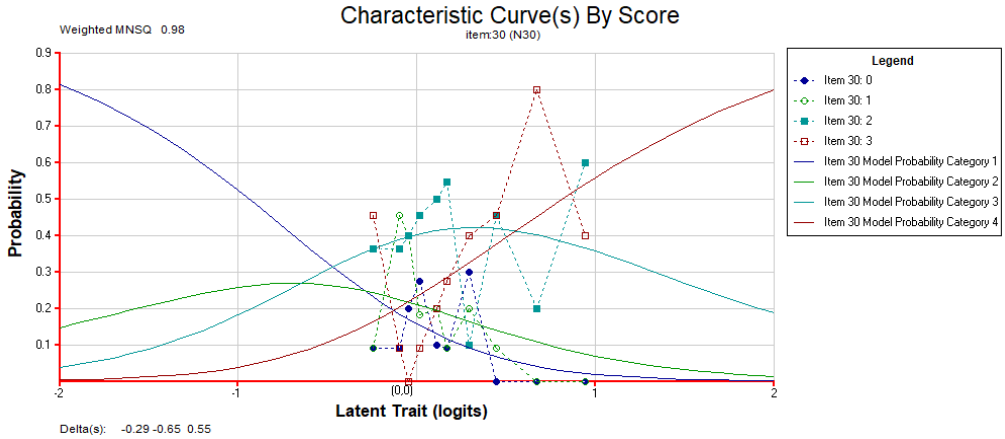


Figura C.30: Curva Característica del ítem N°30



Anexo D

Curvas de Probabilidad Acumulada

Figura D.1: Curva de Probabilidad Acumulada N°1

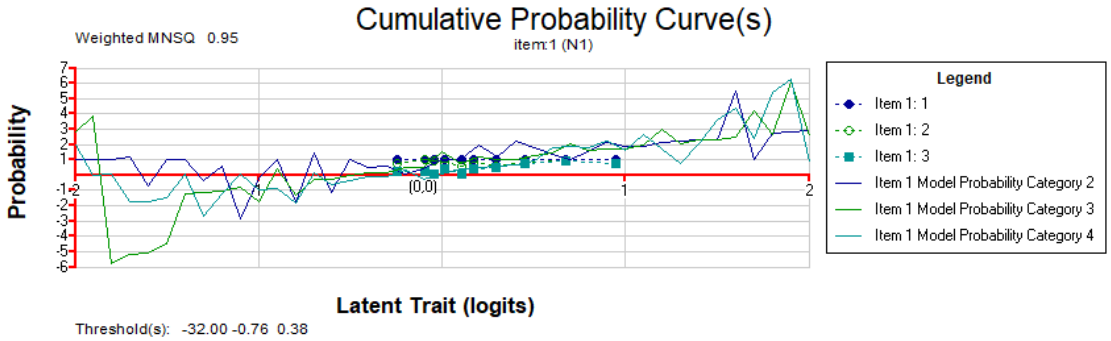


Figura D.2: Curva de Probabilidad Acumulada N°2

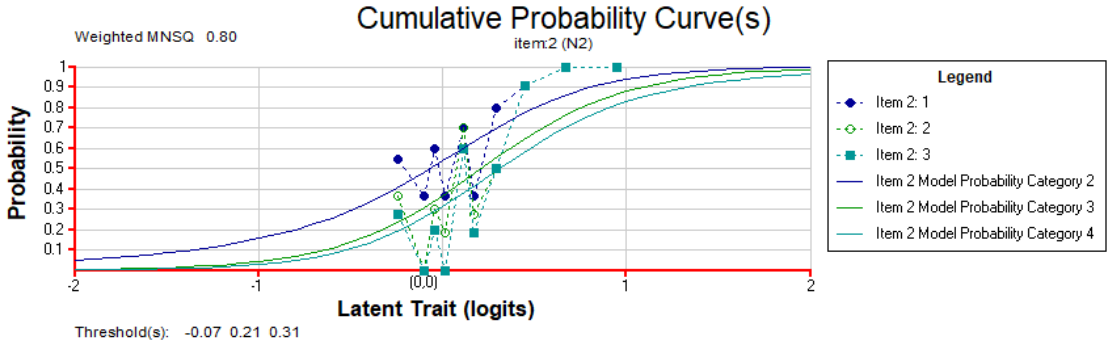


Figura D.3: Curva de Probabilidad Acumulada N°3

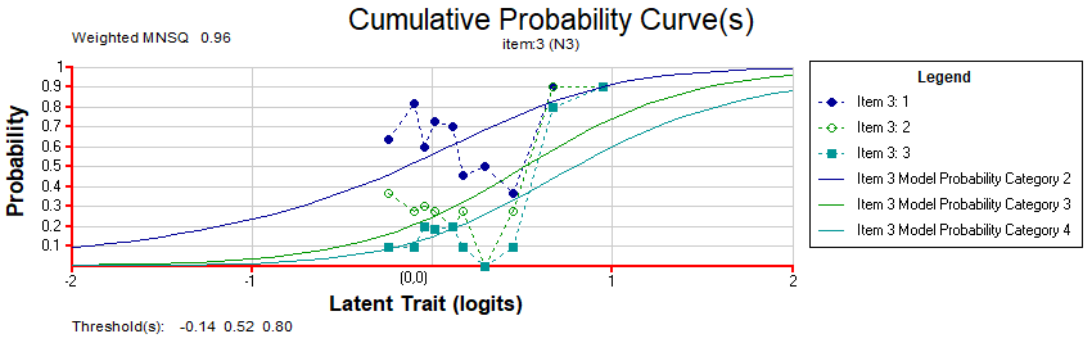


Figura D.4: Curva de Probabilidad Acumulada N°4

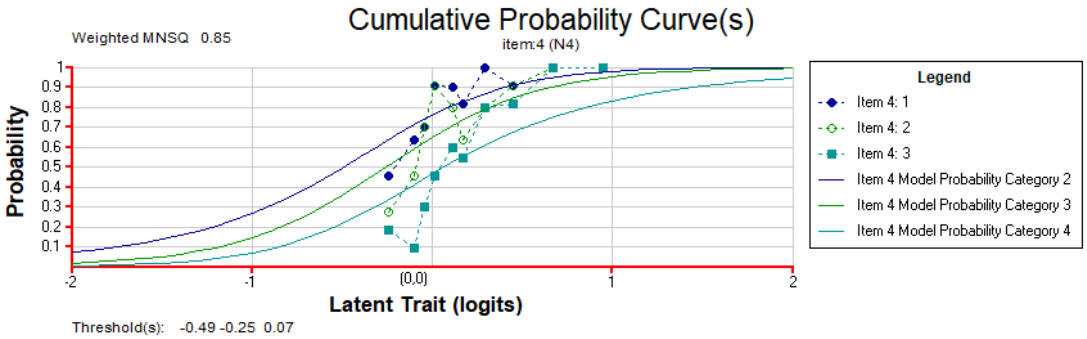


Figura D.5: Curva de Probabilidad Acumulada N°5

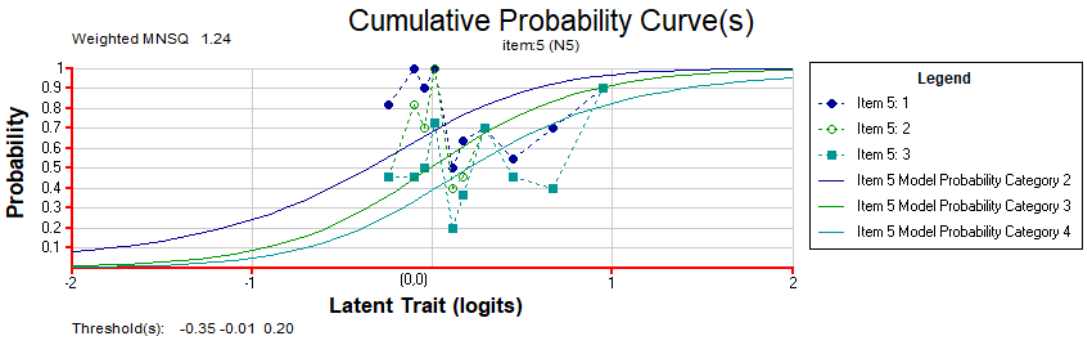


Figura D.6: Curva de Probabilidad Acumulada N°6

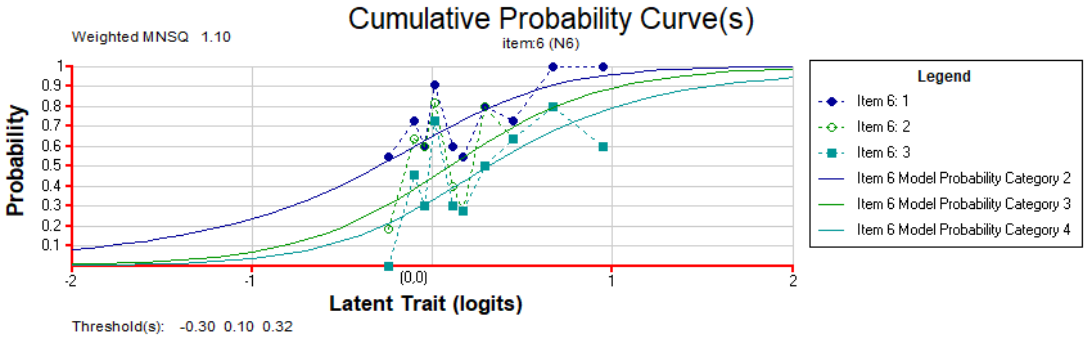


Figura D.7: Curva de Probabilidad Acumulada N°7

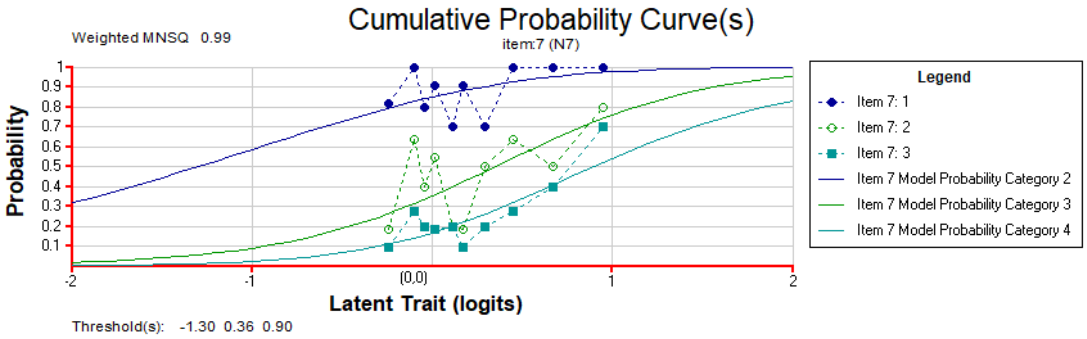


Figura D.8: Curva de Probabilidad Acumulada N°8

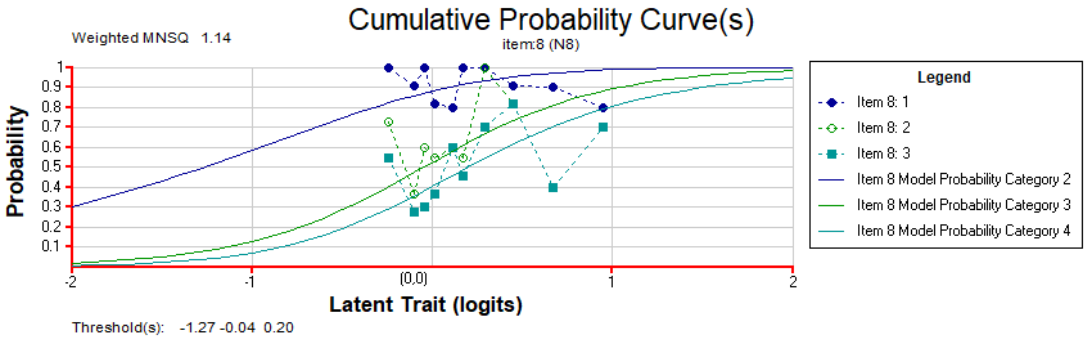


Figura D.9: Curva de Probabilidad Acumulada N°9

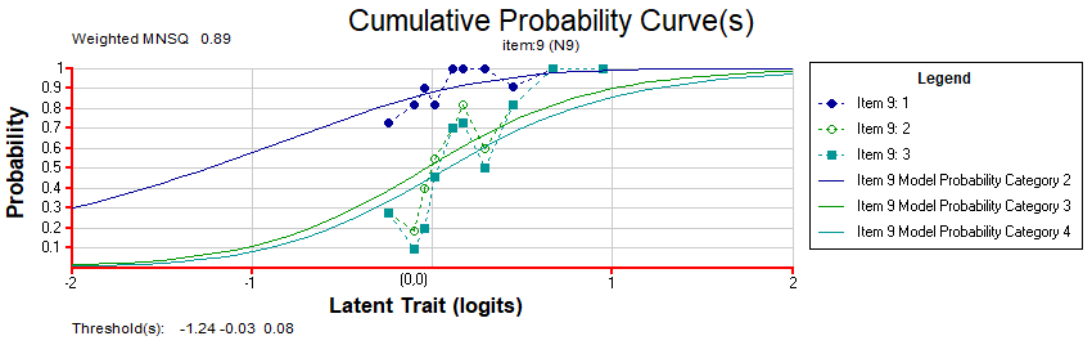


Figura D.10: Curva de Probabilidad Acumulada N°10

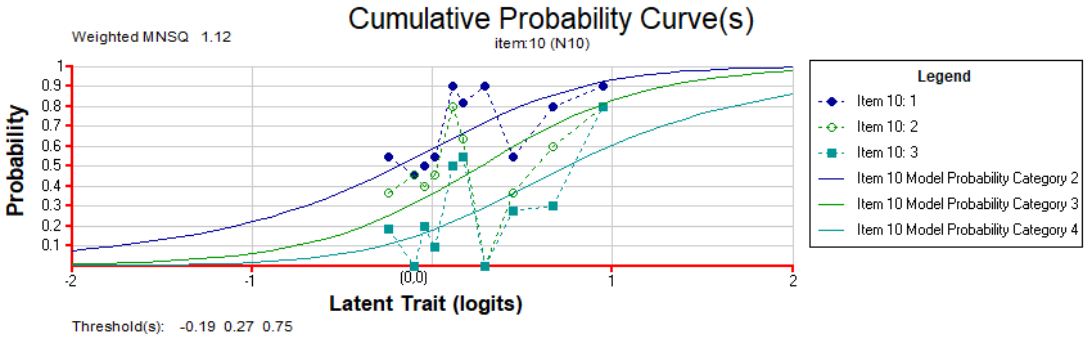


Figura D.11: Curva de Probabilidad Acumulada N°11

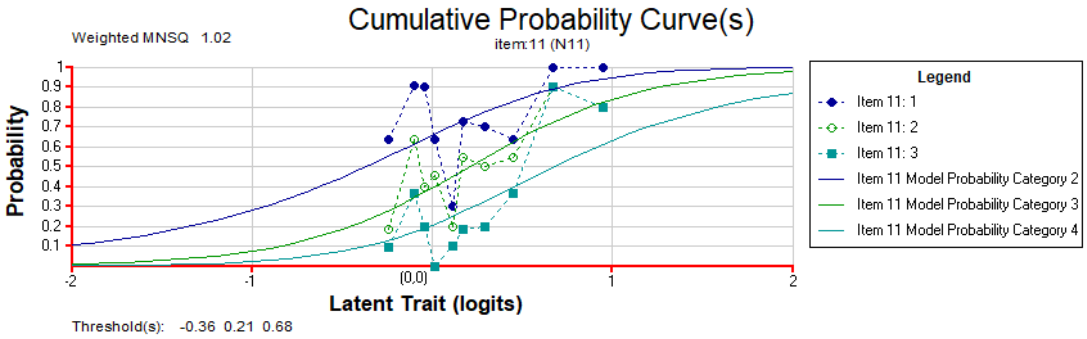


Figura D.12: Curva de Probabilidad Acumulada N°12

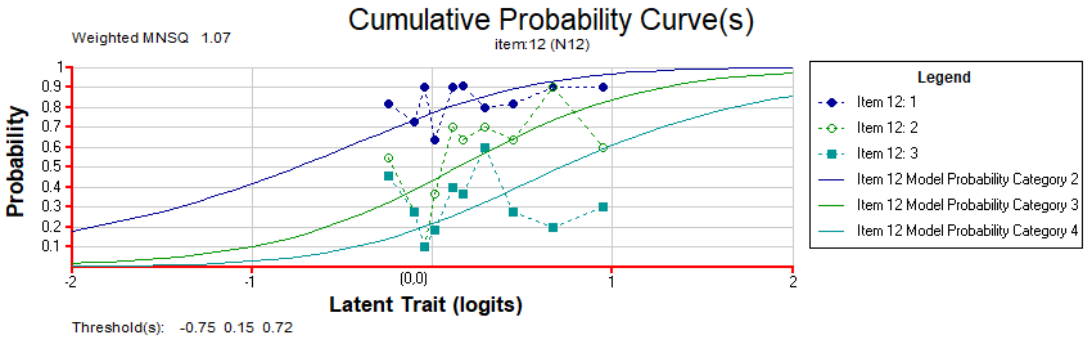


Figura D.13: Curva de Probabilidad Acumulada N°13

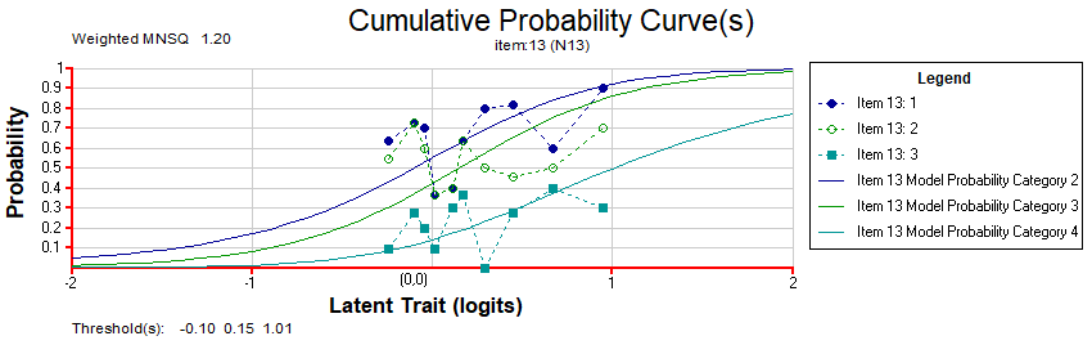


Figura D.14: Curva de Probabilidad Acumulada N°14

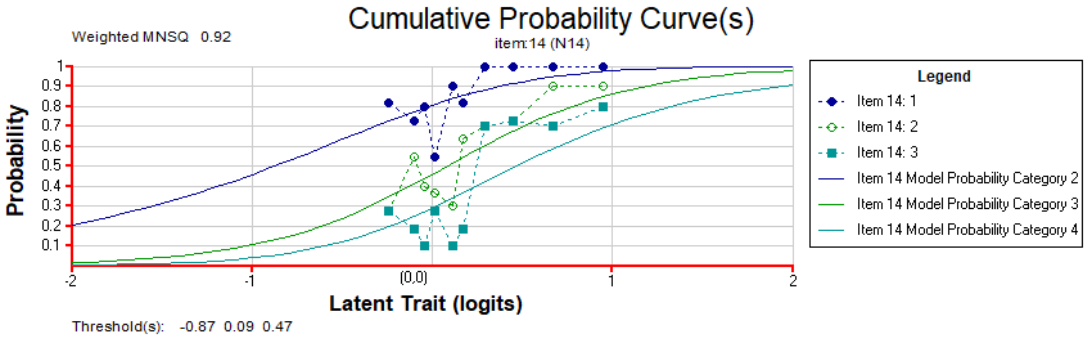


Figura D.15: Curva de Probabilidad Acumulada N°15

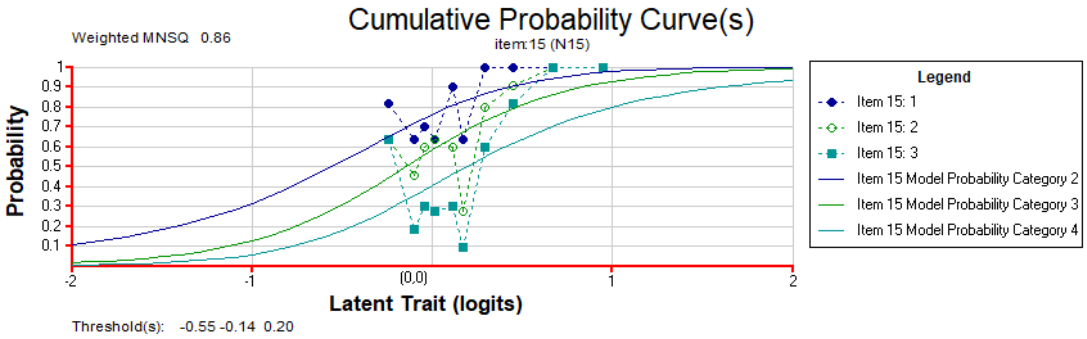


Figura D.16: Curva de Probabilidad Acumulada N°16

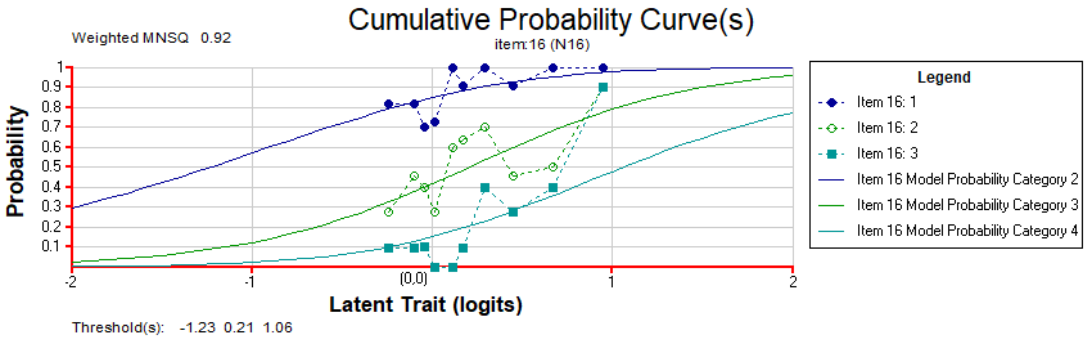


Figura D.17: Curva de Probabilidad Acumulada N°17

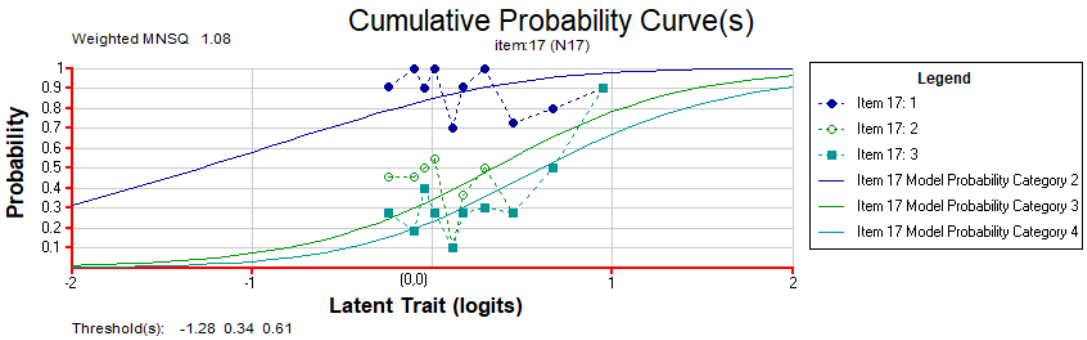


Figura D.18: Curva de Probabilidad Acumulada N°18

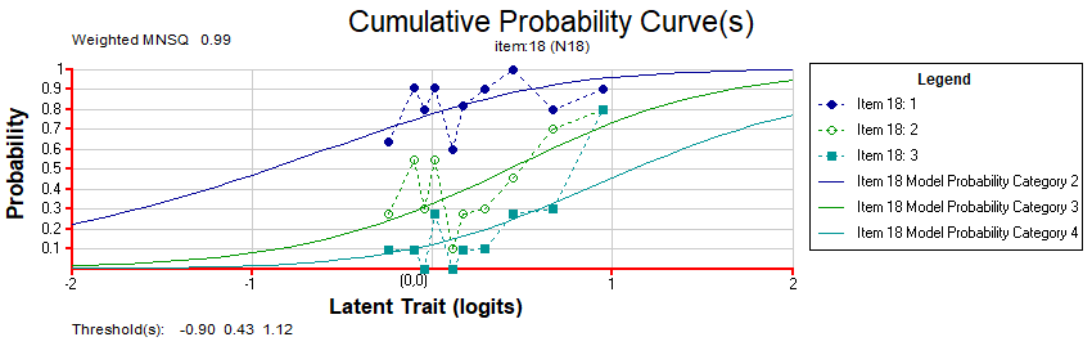


Figura D.19: Curva de Probabilidad Acumulada N°19

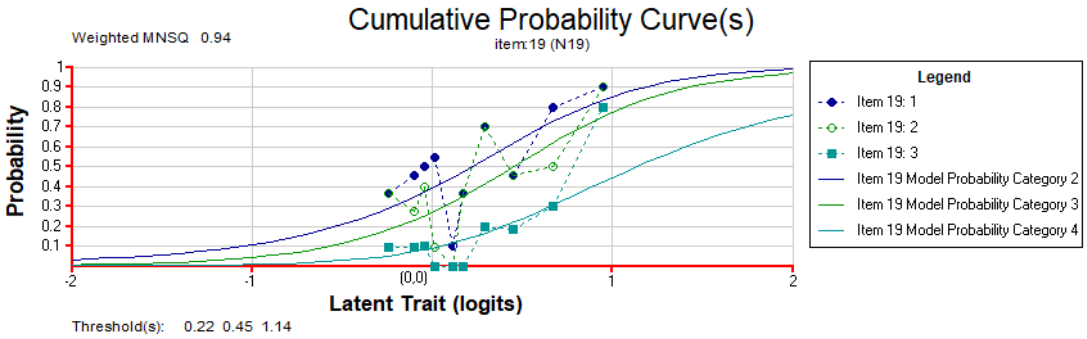


Figura D.20: Curva de Probabilidad Acumulada N°20

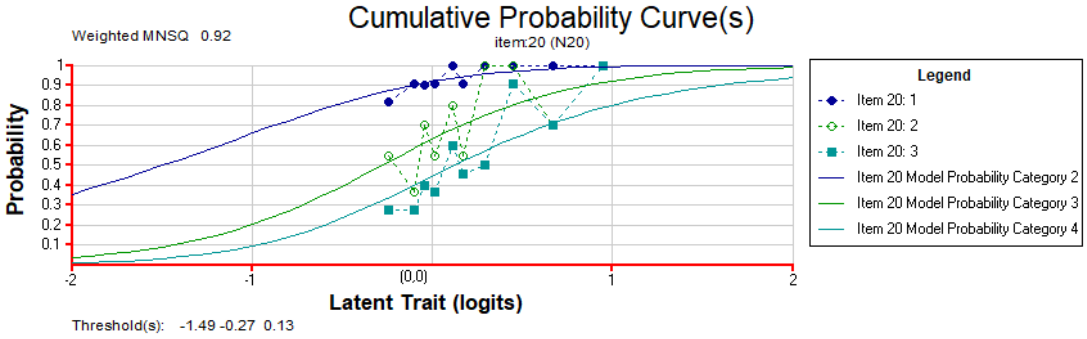


Figura D.21: Curva de Probabilidad Acumulada N°21

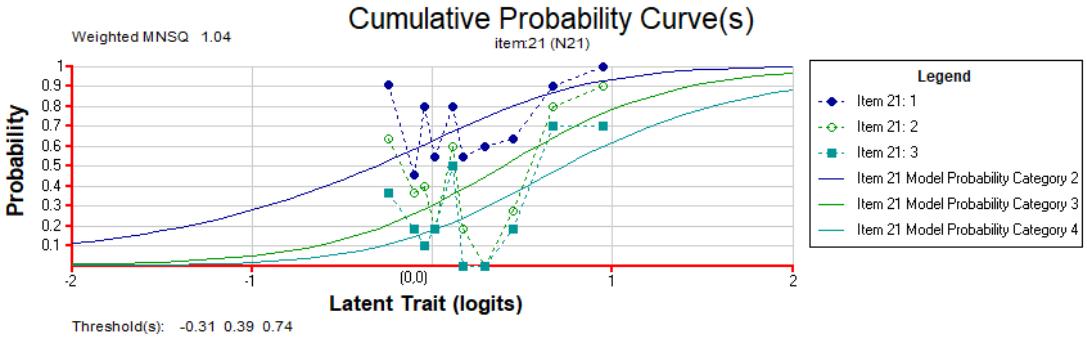


Figura D.22: Curva de Probabilidad Acumulada N°22

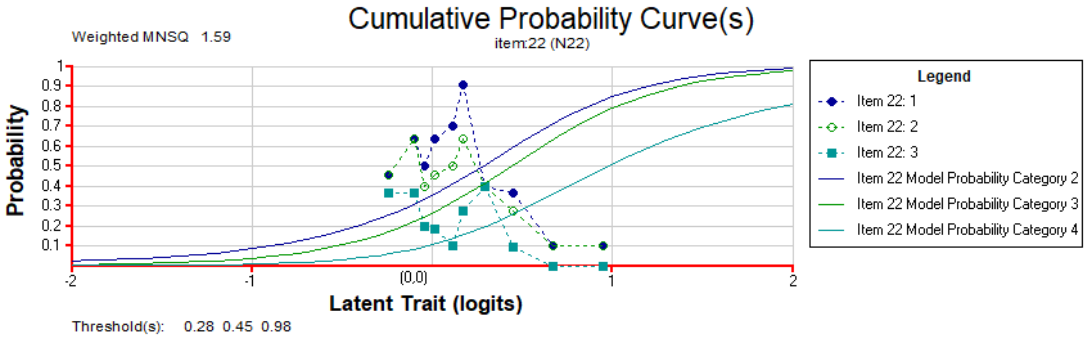


Figura D.23: Curva de Probabilidad Acumulada N°23

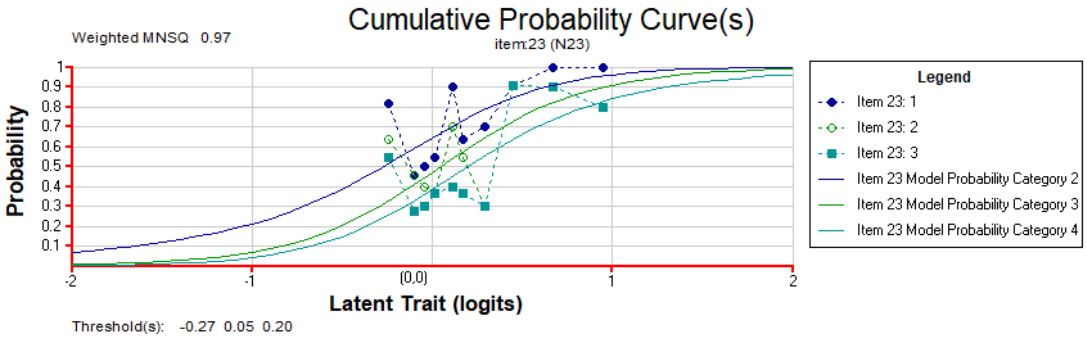


Figura D.24: Curva de Probabilidad Acumulada N°24

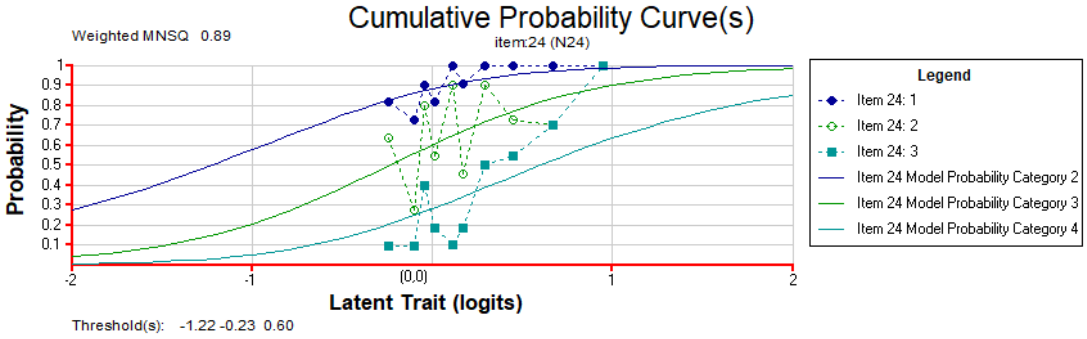


Figura D.25: Curva de Probabilidad Acumulada N°25

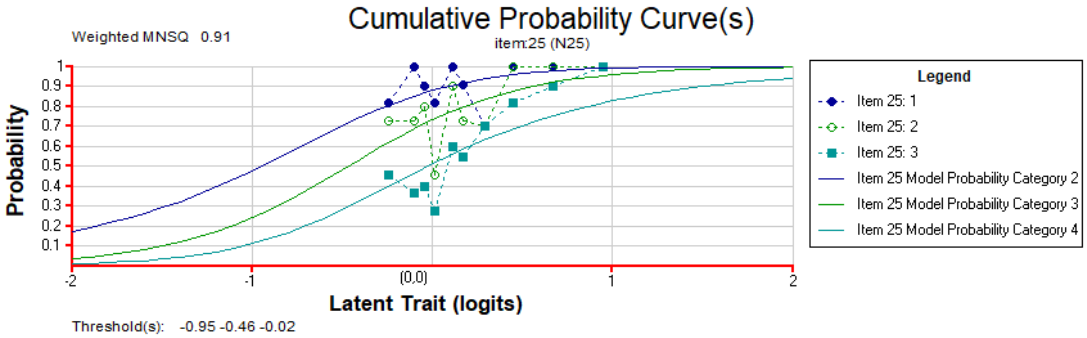


Figura D.26: Curva de Probabilidad Acumulada N°26

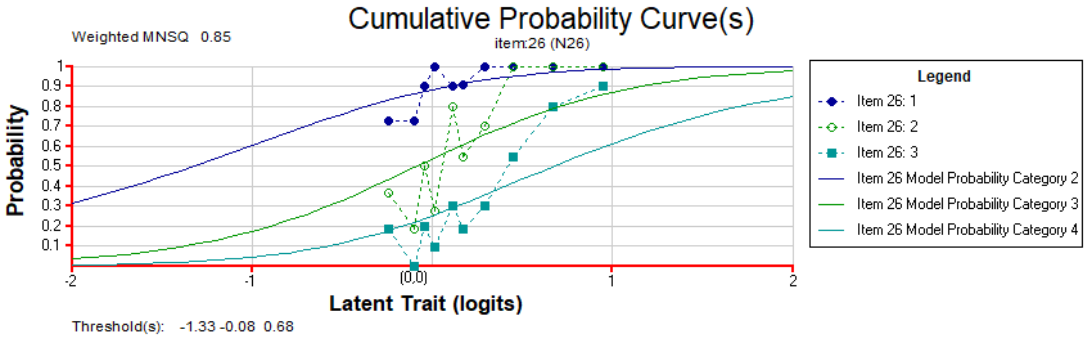


Figura D.27: Curva de Probabilidad Acumulada N°27

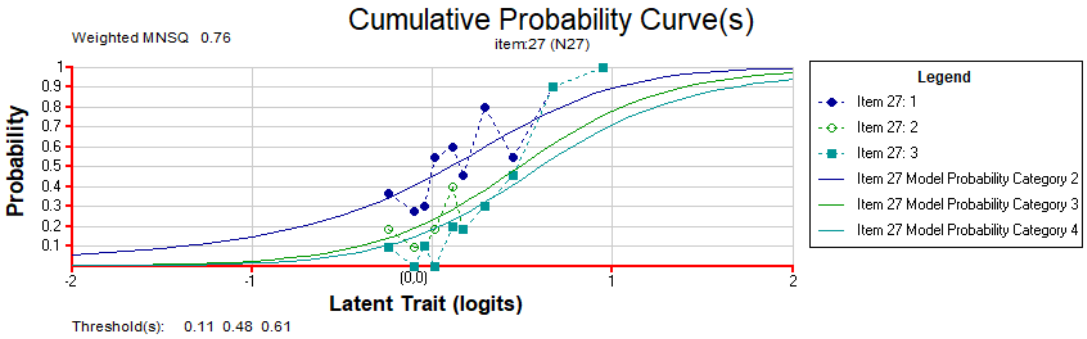


Figura D.28: Curva de Probabilidad Acumulada N°28

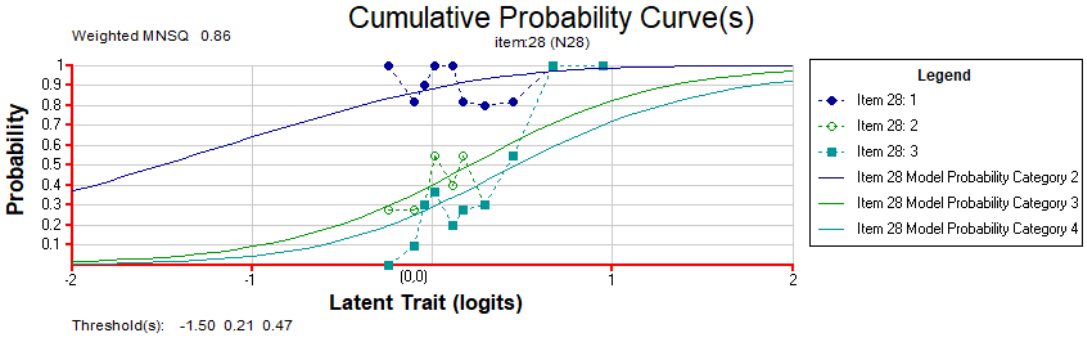


Figura D.29: Curva de Probabilidad Acumulada N°29

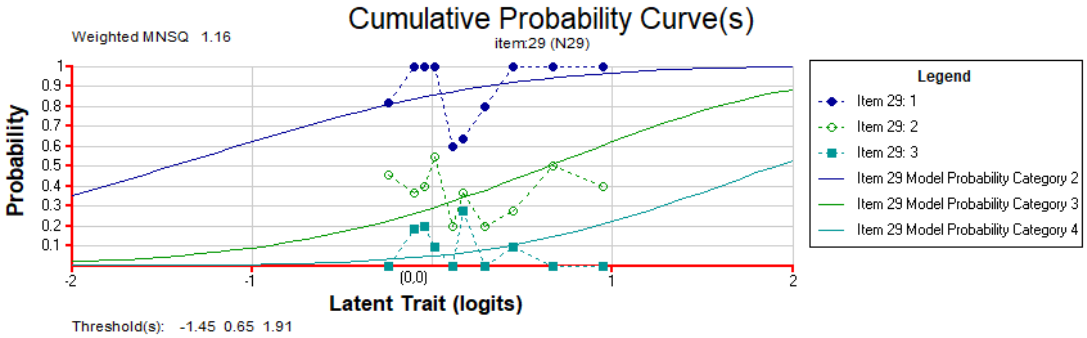
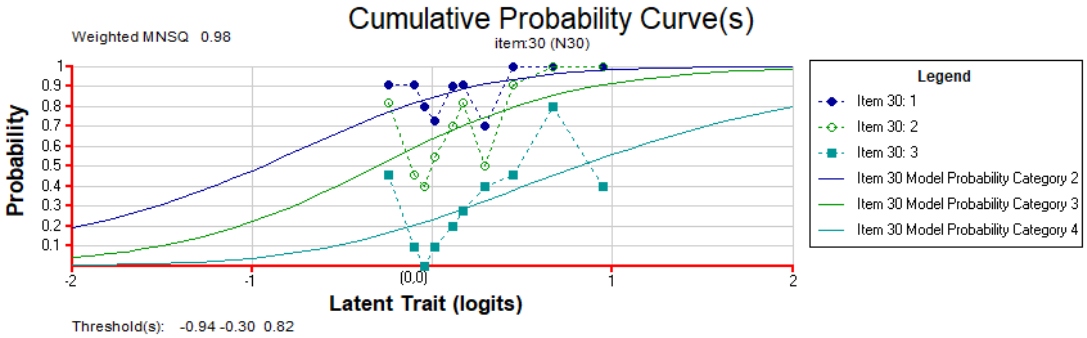


Figura D.30: Curva de Probabilidad Acumulada N°30



Anexo E

Estadísticas Tradicionales del ítem

Figura E.1: Estadísticas Tradicionales del ítem N°01

item:1 (N1)
Cases for this item 105 Discrimination 0.37
Item Threshold(s): -32.00 -0.76 0.38 Weighted MNSQ 0.95
Item Delta(s): -0.19 -0.37 -0.01

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	1.00	21	20.00	-0.06	-0.65(.517)	0.21	0.33
2	2.00	36	34.29	-0.49	-5.68(.000)	-0.01	0.16
3	3.00	48	45.71	0.52	6.12(.000)	0.40	0.37

Figura E.2: Estadísticas Tradicionales del ítem N°02

item:2 (N2)
Cases for this item 105 Discrimination 0.67
Item Threshold(s): -0.07 0.21 0.31 Weighted MNSQ 0.80
Item Delta(s): 1.01 1.13 -1.75

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
0	0.00	36	34.29	-0.44	-5.03(.000)	0.02	0.16
1	1.00	15	14.29	-0.36	-3.95(.000)	-0.01	0.20
2	2.00	6	5.71	-0.08	-0.86(.391)	0.00	0.15
3	3.00	48	45.71	0.72	10.45(.000)	0.47	0.35

Figura E.3: Estadísticas Tradicionales del ítem N°03

```
item:3 (N3)
Cases for this item    105   Discrimination  0.46
Item Threshold(s):    -0.14  0.52  0.80   Weighted MNSQ   0.96
Item Delta(s):      0.35  1.17 -0.43
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       36      34.29   -0.14    -1.49(.140)  0.18     0.28
   1       1.00       30      28.57   -0.32    -3.43(.001)  0.04     0.18
   2       2.00       12      11.43   -0.16    -1.64(.103)  0.07     0.30
   3       3.00       27      25.71    0.60     7.71(.000)  0.54     0.41
================================================================================
```

Figura E.4: Estadísticas Tradicionales del ítem N°04

```
item:4 (N4)
Cases for this item    105   Discrimination  0.56
Item Threshold(s):    -0.49 -0.25  0.07   Weighted MNSQ   0.85
Item Delta(s):      0.79 -0.52 -0.93
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       18      17.14   -0.40    -4.39(.000) -0.06     0.18
   1       1.00        9       8.57   -0.18    -1.81(.073)  0.03     0.24
   2       2.00       18      17.14   -0.26    -2.75(.007)  0.01     0.13
   3       3.00       60      57.14    0.60     7.62(.000)  0.40     0.35
================================================================================
```

Figura E.5: Estadísticas Tradicionales del ítem N°05

```
Item 5 (N5)
Cases for this item    105   Discrimination  0.09
Item Threshold(s):    -0.35 -0.01  0.20   Weighted MNSQ   1.24
Item Delta(s):      0.58  0.42 -1.20
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       24      22.86    0.08     0.81(.422)  0.29     0.28
   1       1.00       15      14.29   -0.22    -2.25(.026)  0.09     0.34
   2       2.00       12      11.43   -0.20    -2.03(.045)  0.01     0.09
   3       3.00       54      51.43    0.21     2.18(.031)  0.27     0.40
================================================================================
```

Figura E.6: Estadísticas Tradicionales del ítem N°06

```
item:6 (N6)
Cases for this item    105   Discrimination  0.27
Item Threshold(s):    -0.30  0.10  0.32   Weighted MNSQ   1.10
Item Delta(s):      0.53  0.61 -1.07
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       27      25.71   -0.26    -2.77(.007)  0.06     0.21
   1       1.00       18      17.14    0.11     1.12(.267)  0.30     0.47
   2       2.00       12      11.43   -0.23    -2.43(.017)  0.00     0.21
   3       3.00       48      45.71    0.30     3.16(.002)  0.33     0.34
================================================================================
```

Figura E.7: Estadísticas Tradicionales del ítem N°07

```
item:7 (N7)
Cases for this item    105   Discrimination  0.36
Item Threshold(s):    -1.30  0.36  0.90   Weighted MNSQ   0.99
Item Delta(s):     -1.18  0.99  0.10
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00       12      11.43   -0.19    -1.95(.053)  0.06     0.20
   1       1.00       45      42.86   -0.16    -1.68(.096)  0.15     0.30
   2       2.00       21      20.00   -0.05    -0.47(.637)  0.19     0.31
   3       3.00       27      25.71    0.37     3.98(.000)  0.42     0.44
===============================================================================
```

Figura E.8: Estadísticas Tradicionales del ítem N°08

```
item:8 (N8)
Cases for this item    105   Discrimination  0.07
Item Threshold(s):    -1.27 -0.04  0.20   Weighted MNSQ   1.14
Item Delta(s):     -1.10  1.10 -1.21
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        9       8.57    0.17     1.79(.077)  0.34     0.39
   1       1.00       30      28.57   -0.14    -1.43(.155)  0.14     0.32
   2       2.00       12      11.43   -0.29    -3.09(.003) -0.00     0.22
   3       3.00       54      51.43    0.21     2.23(.028)  0.29     0.37
===============================================================================
```

Figura E.9: Estadísticas Tradicionales del ítem N°09

```
item:9 (N9)
Cases for this item    105   Discrimination  0.50
Item Threshold(s):    -1.24 -0.03  0.08   Weighted MNSQ   0.89
Item Delta(s):     -1.10  1.79 -2.02
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        9       8.57   -0.14    -1.46(.147) -0.05     0.18
   1       1.00       30      28.57   -0.42    -4.74(.000)  0.00     0.20
   2       2.00        6       5.71   -0.17    -1.80(.075)  0.03     0.12
   3       3.00       60      57.14    0.55     6.66(.000)  0.39     0.36
===============================================================================
```

Figura E.10: Estadísticas Tradicionales del ítem N°10

```
item:10 (N10)
Cases for this item    105   Discrimination  0.29
Item Threshold(s):    -0.19  0.27  0.75   Weighted MNSQ   1.12
Item Delta(s):      0.60  0.24  0.00
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00       33      31.43   -0.23    -2.45(.016)  0.10     0.30
   1       1.00       21      20.00    0.05     0.47(.637)  0.27     0.30
   2       2.00       21      20.00   -0.15    -1.55(.123)  0.09     0.26
   3       3.00       30      28.57    0.33     3.60(.000)  0.41     0.42
===============================================================================
```

Figura E.11: Estadísticas Tradicionales del ítem N°11

```
item:11 (N11)
Cases for this item    105   Discrimination  0.39
Item Threshold(s):    -0.36  0.21  0.68   Weighted MNSQ   1.02
Item Delta(s):      0.26  0.36 -0.11
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       27      25.71   -0.16    -1.67(.099)  0.12     0.22
   1       1.00       24      22.86   -0.18    -1.87(.064)  0.10     0.36
   2       2.00       21      20.00   -0.24    -2.56(.012)  0.09     0.18
   3       3.00       33      31.43    0.53     6.29(.000)  0.47     0.39
================================================================================
```

Figura E.12: Estadísticas Tradicionales del ítem N°12

```
item:12 (N12)
Cases for this item    105   Discrimination  0.22
Item Threshold(s):    -0.75  0.15  0.72   Weighted MNSQ   1.07
Item Delta(s):     -0.38  0.44  0.02
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       18      17.14   -0.13    -1.28(.203)  0.15     0.33
   1       1.00       30      28.57   -0.22    -2.31(.023)  0.11     0.35
   2       2.00       24      22.86    0.25     2.63(.010)  0.43     0.33
   3       3.00       33      31.43    0.09     0.93(.356)  0.21     0.33
================================================================================
```

Figura E.13: Estadísticas Tradicionales del ítem N°13

```
item:13 (N13)
Cases for this item    105   Discrimination  0.11
Item Threshold(s):    -0.10  0.15  1.01   Weighted MNSQ   1.20
Item Delta(s):      1.25 -0.77  0.68
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       36      34.29   -0.17    -1.75(.084)  0.14     0.29
   1       1.00       12      11.43    0.20     2.06(.042)  0.43     0.37
   2       2.00       33      31.43   -0.03    -0.30(.766)  0.17     0.38
   3       3.00       24      22.86    0.07     0.75(.455)  0.29     0.35
================================================================================
```

Figura E.14: Estadísticas Tradicionales del ítem N°14

```
item:14 (N14)
Cases for this item    105   Discrimination  0.49
Item Threshold(s):    -0.87  0.09  0.47   Weighted MNSQ   0.92
Item Delta(s):     -0.57  0.71 -0.53
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       15      14.29   -0.36    -3.87(.000) -0.02     0.14
   1       1.00       30      28.57   -0.13    -1.38(.171)  0.10     0.27
   2       2.00       18      17.14   -0.13    -1.28(.203)  0.16     0.30
   3       3.00       42      40.00    0.48     5.48(.000)  0.42     0.38
================================================================================
```

Figura E.15: Estadísticas Tradicionales del ítem N°15

```
item:15 (N15)
Cases for this item    105   Discrimination  0.56
Item Threshold(s):    -0.55 -0.14  0.20   Weighted MNSQ   0.86
Item Delta(s):      0.29  0.00 -0.81
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       18      17.14   -0.42    -4.64(.000) -0.02     0.14
   1       1.00       15      14.29   -0.19    -1.96(.052)  0.09     0.20
   2       2.00       18      17.14   -0.11    -1.09(.278)  0.07     0.15
   3       3.00       54      51.43    0.53     6.29(.000)  0.38     0.40
================================================================================
```

Figura E.16: Estadísticas Tradicionales del ítem N°16

```
item:16 (N16)
Cases for this item    105   Discrimination  0.48
Item Threshold(s):    -1.23  0.21  1.06   Weighted MNSQ   0.92
Item Delta(s):     -1.04  0.49  0.57
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       12      11.43   -0.30    -3.18(.002) -0.04     0.18
   1       1.00       39      37.14   -0.06    -0.60(.549)  0.17     0.31
   2       2.00       30      28.57   -0.26    -2.71(.008)  0.10     0.19
   3       3.00       24      22.86    0.57     7.09(.000)  0.57     0.40
================================================================================
```

Figura E.17: Estadísticas Tradicionales del ítem N°17

```
item:17 (N17)
Cases for this item    105   Discrimination  0.24
Item Threshold(s):    -1.28  0.34  0.61   Weighted MNSQ   1.08
Item Delta(s):     -1.19  1.54 -0.76
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       12      11.43    0.15     1.51(.134)  0.32     0.31
   1       1.00       45      42.86   -0.35    -3.75(.000)  0.12     0.25
   2       2.00       12      11.43   -0.17    -1.80(.075) -0.01     0.20
   3       3.00       36      34.29    0.38     4.16(.000)  0.38     0.44
================================================================================
```

Figura E.18: Estadísticas Tradicionales del ítem N°18

```
item:18 (N18)
Cases for this item    105   Discrimination  0.36
Item Threshold(s):    -0.90  0.43  1.12   Weighted MNSQ   0.99
Item Delta(s):     -0.70  0.80  0.50
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       18      17.14   -0.09    -0.96(.338)  0.13     0.32
   1       1.00       42      40.00   -0.17    -1.71(.090)  0.15     0.26
   2       2.00       24      22.86   -0.21    -2.23(.028)  0.13     0.30
   3       3.00       21      20.00    0.52     6.14(.000)  0.54     0.43
================================================================================
```

Figura E.19: Estadísticas Tradicionales del ítem N°19

```
item:19 (N19)
Cases for this item    105   Discrimination  0.48
Item Threshold(s):     0.22  0.45  1.14   Weighted MNSQ   0.94
Item Delta(s):      1.63 -0.41  0.70
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       51      48.57   -0.32    -3.45(.001)  0.11     0.27
   1       1.00       12      11.43   -0.08    -0.81(.421)  0.16     0.29
   2       2.00       24      22.86   -0.08    -0.77(.441)  0.19     0.30
   3       3.00       18      17.14    0.58     7.20(.000)  0.60     0.42
================================================================================
```

Figura E.20: Estadísticas Tradicionales del ítem N°20

```
item:20 (N20)
Cases for this item    105   Discrimination  0.47
Item Threshold(s):    -1.49 -0.27  0.13   Weighted MNSQ   0.92
Item Delta(s):     -1.29  0.46 -0.88
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        6       5.71   -0.20    -2.12(.036) -0.09     0.15
   1       1.00       24      22.86   -0.26    -2.78(.006)  0.07     0.25
   2       2.00       18      17.14   -0.26    -2.75(.007)  0.06     0.24
   3       3.00       57      54.29    0.52     6.11(.000)  0.36     0.37
================================================================================
```

Figura E.21: Estadísticas Tradicionales del ítem N°21

```
item:21 (N21)
Cases for this item    105   Discrimination  0.33
Item Threshold(s):    -0.31  0.39  0.74   Weighted MNSQ   1.04
Item Delta(s):      0.15  0.94 -0.33
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       30      28.57   -0.25    -2.66(.009)  0.12     0.22
   1       1.00       30      28.57   -0.06    -0.58(.561)  0.17     0.28
   2       2.00       15      14.29    0.00     0.03(.977)  0.18     0.41
   3       3.00       30      28.57    0.31     3.29(.001)  0.39     0.44
================================================================================
```

Figura E.22: Estadísticas Tradicionales del ítem N°22

```
item:22 (N22)
Cases for this item    105   Discrimination -0.30
Item Threshold(s):     0.28  0.45  0.98   Weighted MNSQ   1.59
Item Delta(s):      1.97 -0.57  0.41
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       54      51.43    0.31     3.29(.001)  0.33     0.40
   1       1.00        9       8.57   -0.12    -1.20(.232)  0.14     0.11
   2       2.00       21      20.00   -0.08    -0.83(.409)  0.14     0.28
   3       3.00       21      20.00   -0.22    -2.30(.023)  0.04     0.23
================================================================================
```

Figura E.23: Estadísticas Tradicionales del ítem N°23

```
item:23 (N23)
Cases for this item    105   Discrimination  0.43
Item Threshold(s):    -0.27  0.05  0.20   Weighted MNSQ   0.97
Item Delta(s):      0.70  0.71 -1.49
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00       27      25.71   -0.39    -4.24(.000)  0.03     0.18
   1       1.00       15      14.29    0.01     0.10(.923)  0.23     0.36
   2       2.00        9       8.57   -0.19    -1.99(.049)  0.02     0.12
   3       3.00       54      51.43    0.44     4.95(.000)  0.34     0.39
===============================================================================
```

Figura E.24: Estadísticas Tradicionales del ítem N°24

```
item:24 (N24)
Cases for this item    105   Discrimination  0.52
Item Threshold(s):    -1.22 -0.23  0.60   Weighted MNSQ   0.89
Item Delta(s):     -0.87 -0.13  0.14
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        9       8.57   -0.30    -3.20(.002) -0.07     0.12
   1       1.00       24      22.86   -0.09    -0.89(.377)  0.15     0.28
   2       2.00       33      31.43   -0.38    -4.17(.000)  0.05     0.20
   3       3.00       39      37.14    0.62     7.92(.000)  0.47     0.38
===============================================================================
```

Figura E.25: Estadísticas Tradicionales del ítem N°25

```
item:25 (N25)
Cases for this item    105   Discrimination  0.47
Item Threshold(s):    -0.95 -0.46 -0.02   Weighted MNSQ   0.91
Item Delta(s):     -0.20 -0.40 -0.84
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        9       8.57   -0.18    -1.81(.073)  0.04     0.24
   1       1.00       12      11.43   -0.34    -3.70(.000) -0.02     0.13
   2       2.00       21      20.00   -0.23    -2.36(.020)  0.06     0.22
   3       3.00       63      60.00    0.51     5.99(.000)  0.35     0.38
===============================================================================
```

Figura E.26: Estadísticas Tradicionales del ítem N°26

```
item:26 (N26)
Cases for this item    105   Discrimination  0.61
Item Threshold(s):    -1.33 -0.08  0.68   Weighted MNSQ   0.85
Item Delta(s):     -1.09  0.20  0.13
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        9       8.57   -0.22    -2.26(.026) -0.09     0.15
   1       1.00       30      28.57   -0.41    -4.53(.000) -0.00     0.15
   2       2.00       30      28.57   -0.09    -0.95(.343)  0.21     0.27
   3       3.00       36      34.29    0.61     7.71(.000)  0.49     0.38
===============================================================================
```

Figura E.27: Estadísticas Tradicionales del ítem N°27

```
item:27 (N27)
Cases for this item    105   Discrimination  0.71
Item Threshold(s):     0.11  0.48  0.61   Weighted MNSQ   0.76
Item Delta(s):      0.93  1.51 -1.33
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       45      42.86   -0.46    -5.25(.000)  0.04     0.22
   1       1.00       21      20.00   -0.27    -2.89(.005)  0.09     0.21
   2       2.00        6       5.71   -0.10    -1.07(.288) -0.01     0.10
   3       3.00       33      31.43    0.78    12.56(.000)  0.58     0.33
================================================================================
```

Figura E.28: Estadísticas Tradicionales del ítem N°28

```
item:28 (N28)
Cases for this item    105   Discrimination  0.55
Item Threshold(s):    -1.50  0.21  0.47   Weighted MNSQ   0.86
Item Delta(s):     -1.42  1.46 -0.93
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        9       8.57   -0.08    -0.77(.441)  0.17     0.21
   1       1.00       42      40.00   -0.46    -5.20(.000)  0.02     0.21
   2       2.00       12      11.43   -0.21    -2.19(.031) -0.01     0.15
   3       3.00       42      40.00    0.64     8.38(.000)  0.49     0.36
================================================================================
```

Figura E.29: Estadísticas Tradicionales del ítem N°29

```
item:29 (N29)
Cases for this item    105   Discrimination  0.01
Item Threshold(s):    -1.45  0.65  1.91   Weighted MNSQ   1.16
Item Delta(s):     -1.34  0.85  1.60
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       12      11.43   -0.15    -1.49(.139)  0.10     0.16
   1       1.00       54      51.43    0.11     1.10(.274)  0.25     0.36
   2       2.00       30      28.57    0.06     0.57(.571)  0.26     0.42
   3       3.00        9       8.57   -0.12    -1.20(.232)  0.07     0.16
================================================================================
```

Figura E.30: Estadísticas Tradicionales del ítem N°30

```
item:30 (N30)
Cases for this item    105   Discrimination  0.37
Item Threshold(s):    -0.94 -0.30  0.82   Weighted MNSQ   0.98
Item Delta(s):     -0.29 -0.65  0.55
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       12      11.43   -0.25    -2.59(.011)  0.05     0.17
   1       1.00       18      17.14   -0.18    -1.87(.065)  0.04     0.18
   2       2.00       42      40.00    0.04     0.39(.700)  0.24     0.39
   3       3.00       33      31.43    0.28     2.92(.004)  0.36     0.36
================================================================================
```

Figura E.31: Otras Estadísticas Tradicionales

The following traditional statistics are only meaningful for complete designs and when the amount of missing data is minimal.
In this analysis 0.00% of the data are missing.

The following results are scaled to assume that a single response was provided for each item.

N	105
Mean	51.71
Standard Deviation	12.34
Variance	152.17
Skewness	0.95
Kurtosis	-0.18
Standard error of mean	1.20
Standard error of measurement	5.76
Coefficient Alpha	0.78

Anexo F

Dificultades de Ítem y Step (Tipos de error)

Figura F.1: Relación entre dificultad y probabilidad de cada step

		Probabilidades de cada tipo de error			
N°	Estimate	Step0	Step1	Step2	Step3
20	-0,57	5,71	22,86	17,14	54,29
25	-0,479	8,57	11,43	20	60
9	-0,445	8,57	28,57	5,71	57,14
8	-0,404	8,57	28,57	11,43	51,43
28	-0,299	8,57	40	11,43	40
24	-0,287	8,57	22,86	31,43	37,14
26	-0,252	8,57	28,57	28,57	34,29
4	-0,218	17,14	8,57	17,14	57,14
1	-0,191	0	20	34,29	45,71
15	-0,172	17,14	14,29	17,14	51,43
17	-0,137	11,43	42,86	11,43	34,29
14	-0,129	14,29	28,57	17,14	40
30	-0,129	11,43	17,14	40	31,43
5	-0,068	22,86	14,29	11,43	51,43
7	-0,03	11,43	42,86	20	25,71
23	-0,024	25,71	14,29	8,57	51,43
16	0,007	11,43	37,14	28,57	22,86
6	0,024	25,71	17,14	11,43	45,71
12	0,027	17,14	28,57	22,86	31,43
2	0,131	34,29	14,29	5,71	45,71
11	0,171	25,71	22,86	20	31,43
18	0,204	17,14	40	22,86	20
21	0,253	28,57	28,57	14,29	28,57
10	0,281	31,43	20	20	28,57
3	0,364	34,29	28,57	11,43	25,71
27	0,37	42,86	20	5,71	31,43
29	0,37	11,43	51,43	28,57	8,57
13	0,389	34,29	11,43	31,43	22,86
22	0,604	51,43	8,57	20	20
19	0,638	48,57	11,43	22,86	17,14
Promedios		20,1	24,19	18,95	36,76
R Pearson		0,84	-0,09	0,06	-0,78

Anexo G

Ítems del Instrumento

Figura G.1: Ítem N° 1

¿Cuál de las siguientes es una forma equivalente a la expresión $\frac{p+q}{p-q} : \frac{p-q}{p+q}$?

a) 0

b) 1

c) $\frac{(p-q)^2}{(p+q)^2}$

d) $\frac{(p+q)^2}{(p-q)^2}$

e) $\frac{2p+2q}{2p-2q}$

Figura G.2: Ítem N° 2

¿Cuál es el resultado de la siguiente operación?

$$\frac{6x}{3} + \frac{6x}{2}$$

a) 5

b) $5x$

c) $6x$

d) $\frac{6x}{5}$

e) $\frac{12x}{5}$

Figura G.3: Ítem N° 3

La temperatura de un refrigerador recién enchufado puede modelarse a través de la siguiente función

$$f(x) = -2(x - 1) + 14$$

Donde $\boldsymbol{x}$ son las horas en funcionamiento.

En base a esta información, ¿Cuál fue la temperatura cuando se enchufó el refrigerador?

a) 8°

b) 11°

c) 12°

d) 14°

e) 16°

Figura G.4: Ítem N° 4

En la siguiente ecuación

$$3(x + 2) + 4x + 7 = 5(2x - 1) + 4 - x$$

¿Cuál debería ser el valor de x, para que se cumpla la igualdad?

a) $x = 3$

b) $x = 6$

c) $x = 7$

d) $x = \frac{3}{4}$

e) $x = \frac{21}{2}$

Figura G.5: Ítem N° 5

El diagrama muestra una caja de jugo con sus respectivas medidas.

Si su capacidad máxima es de 180 cc ¿Cuánto mide la altura de la cajita?

a) 8 cm

b) 12 cm

c) 27 cm

d) 36 cm

e) 165 cm

Figura G.6: Ítem N° 6

Observe la siguiente expresión algebraica

$$7x^4 - 5y^2 + 2x^2 + xy$$

¿Cuántos términos algebraicos posee?

a) 1

b) 2

c) 3

d) 4

e) 5

Figura G.7: Ítem N° 7

Si se define la operación

$$a \blacksquare b = \frac{a-b}{2}$$

¿Cuál es el resultado de $(3a+4b)\blacksquare(a-2b)$?

a) $4ab$
b) $a+b$
c) $2a+b$
d) $a+3b$
e) $2a+6b$

Figura G.8: Ítem N° 8

De la expresión

$$x(x+y)-y(x-y)$$

¿Cuál de las siguientes alternativas la presenta en su forma reducida?

a) x^2+y^2
b) x^2-y^2
c) $x^2+2y-yx$
d) x^2-yx
e) $(x-y)(x+y)$

Figura G.9: Ítem N° 9

¿Cuál es el producto de los binomios $(4x-3y)$ y $(2y+5x)$?

a) $-7xy$
b) $9x-y$
c) $9x-6y^2$
d) $9x^2+8xy-y$
e) $20x^2-7xy-6y^2$

Figura G.10: Ítem N° 10

De la siguiente resta de fracciones algebraicas

$$\frac{15x-10y}{5z}-\frac{15y-10x}{5z}$$

¿A qué fracción algebraica es equivalente?

a) 0

b) $\frac{x-y}{5z}$

c) $\frac{x-5y}{z}$

d) $\frac{5x-5y}{z}$

e) $\frac{5x-25y}{z}$

Figura G.11: Ítem N° 11

Si $x \neq y$, entonces, ¿Cuál es la forma más reducida de la siguiente expresión?

$$\frac{x^2-y^2}{x-y}+\frac{(x-y)^2}{x-y}-(x-y)$$

a) $2x$

b) $x-y$

c) $x+y$

d) $2x+2y$

e) $x+3y$

Figura G.12: Ítem N° 12

Si se reduce la siguiente expresión algebraica, considerando que $xy \neq 0$

$$\frac{8xy}{4x^3y - 4xy}$$

¿Cuál sería una forma equivalente a dicha fracción?

a) $\frac{1}{x^2y^2}$

b) $\frac{1}{2x^2y^2}$

c) $\frac{2}{x^2-y}$

d) $\frac{2}{(x-y)^2}$

e) $\frac{2}{(x-1)(x+1)}$

Figura G.13: Ítem N° 13

¿Cuál de las siguientes alternativas es una expresión equivalente a $\frac{a-b}{a^3-b^3-(a-b)^3}$?

a) 0

b) $\frac{1}{3ab}$

c) $\frac{a-b}{3ab}$

d) $\frac{a-b}{-b^3}$

e) $\frac{1}{2a^2-ab+2b^2}$

Figura G.14: Ítem N° 14

¿Para qué valor de "x" la expresión $5(x-3) - 2(2x-6)$ es igual a cero?

a) -3

b) 3

c) 7

d) 9

e) 27

Figura G.15: Ítem N° 15

Del enunciado:

"La diferencia entre el quíntuple del sucesor de a y el antecesor del cuadrado del doble de b"

¿Cuál es su representación en lenguaje algebraico?

a) $5a + 1 - [2b^2 - 1]$

b) $5(a + 1) - (2b - 1)^2$

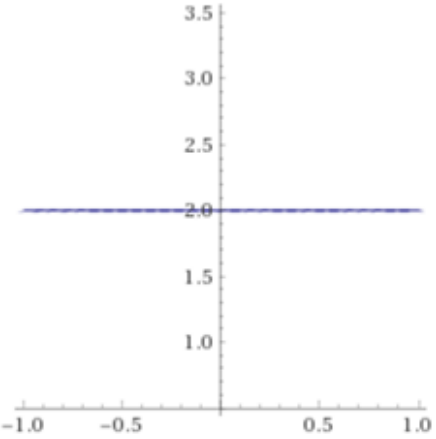
c) $5(a + 1) - [(2b)^2 - 1]$

d) $-5(a + 1) + [(2b)^2 - 1]$

e) $-5 + (a + 1) + [(2b)^2 - 1]$

Figura G.16: Ítem N° 16

De la siguiente ecuación de la recta graficada en un plano cartesiano,



¿Cuál corresponde a su forma algebraica?

a) $x - 2 = 0$

b) $x + 2 = 0$

c) $y - 2 = 0$

d) $y + 2 = 0$

e) $y - 2x = 0$

Figura G.17: Ítem N° 17

De la ecuación lineal $9x - 3y - 6 = 0$ ¿Cuál es su pendiente?

a) -3

b) -2

c) 2

d) 3

e) 9

Figura G.18: Ítem N° 18

Considerando la ecuación lineal

$$5y = 15x + 20$$

¿Cuál es una recta paralela a ella?

a) $y = 15x + 20$

b) $2y = 10x + 8$

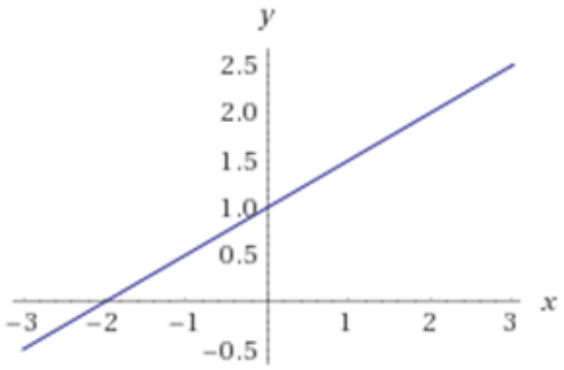
c) $3y = 15x + 20$

d) $4y = 12x + 20$

e) $5y = 20x + 15$

Figura G.19: Ítem N° 19

El gráfico muestra la sección de una recta, entre los valores −3 y +3 del eje X.



¿Cuál es la ecuación que describe el comportamiento de la recta?

a) $y - x = 1$
b) $y - 2x = 1$
c) $2y - x = 1$
d) $2y - x = 2$
e) $2y - x = 4$

Figura G.20: Ítem N° 20

Al resolver la ecuación

$$3x + 14 = 4(2x + 1)$$

¿Cuál es el valor de x?

a) -9
b) -2
c) $\frac{13}{5}$
d) 2
e) 3

Figura G.21: Ítem N° 21

El radio ecuatorial de Saturno mide 60.330 km y es equivalente a 9 veces el radio ecuatorial de la tierra aumentado en 2.928 km.

¿Cuál de las siguientes expresiones muestra la relación entre los radios de la Tierra (T) y Saturno?

a) $9t + 2928 = 60330$

b) $9T - 2928 = 60330$

c) $9T = 60330 + 2928$

d) $9T + 26352 = 60330$

e) $9(t + 2928) = 60330$

Figura G.22: Ítem N° 22

El costo de un automóvil nuevo es de $3.500.000, pero cada año su valor disminuye $200.000 por la depreciación.

¿Cuál es el valor del automóvil "y" después de "x" años?

a) $y = 200.000x + 3.500.000$

b) $y = 200.000x - 3.500.000$

c) $y = -200.000 + 3.500.000x$

d) $y = -200.000x + 3.500.000$

e) $y = (3.500.000 - 200.000)\, x$

Figura G.23: Ítem N° 23

Si se sabe que $a = -1$ y $b = -2$

¿Cuál es el valor de $a - ab$?

a) -3

b) -1

c) 1

d) 2

e) 3

Figura G.24: Ítem N° 24

¿A qué expresión se podría reducir $2a - a - \frac{a}{2}$?

a) 0

b) 1

c) $\frac{1}{2}$

d) $\frac{a}{2}$

e) $\frac{3a}{2}$

Figura G.25: Ítem N° 25

De la expresión *"el doble del cuadrado de a"*,

¿Cuál corresponde a su forma algebraica?

a) $2a^2$

b) $(2a)^2$

c) $2(2a)$

d) 2^2a

e) $2 + a^2$

Figura G.26: Ítem N° 26

Si se reduce a un solo término algebraico, la siguiente expresión

$$\frac{1}{2}a + \frac{2}{3}a - \frac{1}{3}a - a$$

¿Cuál sería la expresión resultante?

a) 0

b) $-\frac{a}{6}$

c) $\frac{a}{6}$

d) $\frac{4a}{6}$

e) $\frac{5a}{6}$

Figura G.27: Ítem N° 27

¿Cuál es una forma equivalente a la siguiente expresión?

$$\frac{a}{a-b} - \frac{b}{a-b} =$$

a) 0

b) 1

c) $\frac{1}{a-b}$

d) $\frac{a+b}{a-b}$

e) $\frac{a-b}{a+b}$

Figura G.28: Ítem N° 28

Un excursionista ha estimado que las horas que demora en subir una colina está dado por la siguiente expresión $T(x) = \frac{x}{600}$, donde x es la altura en metros y T es el tiempo en horas.

¿Cuál es la altura de una colina si demoró 2 horas en subirla?

a) $30m$
b) $120m$
c) $300m$
d) $600m$
e) $1200m$

Figura G.29: Ítem N° 29

¿Para qué valor de x el gráfico de la función $f(x) = -\frac{3}{8}x + 6$ se intersecta con el eje Y?

a) -16
b) -8
c) 0
d) 6
e) 16

Figura G.30: Ítem N° 30

La ganancia semanal (p) de una pista de patinaje, puede calcularse mediante la función

$$f(n) = 8000n - 60000$$

Donde n, es el número de patinadores por semana.

Si en una semana se ganó \$1.020.000.-, ¿Cuántos patinadores asistieron esa semana?

a) 120
b) 127
c) 135
d) 170
e) 202