



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación

TESIS



Estudio del proceso de respuesta a cuestionarios de evaluación estudiantil de la enseñanza desde la perspectiva de las teorías cognitivas de la motivación

Roberto Cáceres

Julio, 2018

ISSN: 2393-7378



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Roberto Cáceres

Estudio del proceso de respuesta a cuestionarios de evaluación estudiantil de la enseñanza desde la perspectiva de las teorías cognitivas de la motivación

Universidad de la República
Área Social

Comisión Sectorial de Enseñanza de la Universidad de la República

Tesis presentada con el objetivo de obtener el título de Magister en Enseñanza Universitaria en el marco del Programa de Especialización y Maestría en Enseñanza Universitaria del Área Social y de la Comisión Sectorial de Enseñanza de la Universidad de la República

Tutor/es: Profesores Bettina Steren dos Santos y Mario Luzardo

Montevideo, 30 de julio 2018

Foto de portada: www.freepik.com



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Dedico, en primer lugar, este trabajo a Silvia y Agustín. También, a todos los estudiantes y docentes que voluntariamente han participado de actividades institucionales de evaluación educativa. A todos aquellos que en su actividad académica cotidiana se comprometen con entusiasmo con una mejora continua de la calidad de la enseñanza y aprendizaje a nivel universitario.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Agradecimientos

Quisiera agradecer a mis orientadores, Bettina Steren y Mario Luzardo, por su apoyo, incentivo y valiosa orientación durante el desarrollo de esta tesis. Pero también, a Mario Luzardo por haber contribuido a mi fuerte interés por modelos de medidas invariantes y modelos de teoría de respuesta a los ítems; y, a Bettina Steren por contribuir a mi profundo interés por los procesos motivacionales en contextos educativos y teorías socio-cognitivas de la motivación. Ambos, con su importante labor académica significan para mí un ejemplo a seguir. También, a Mercedes Collazo, por su esfuerzo e importante labor como Coordinadora Académica del Programa de Especialización y Maestría en Enseñanza Universitaria, que significa un gran apoyo para todos los estudiantes de este posgrado.

Muy especialmente, quiero agradecer a Silvia por todo su amor y apoyo, y además, por su incentivo y paciencia durante el desarrollo de esta tesis.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

(Epígrafe:) *Becoming is better
than being.*

Carol S. Dweck



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

RESUMEN

Los cuestionarios estudiantiles de evaluación de la enseñanza (SET) son ampliamente usados en universidades de todo el mundo como una fuente de retroalimentación acerca de la efectividad o calidad en la enseñanza. Los SET son usados como fuente de datos para mejorar la enseñanza, para tomar decisiones de gestión y administrativas, y para estudiar la relación entre enseñanza y aprendizaje. Si bien se han realizado algunos estudios sobre las actitudes de los estudiantes acerca de los SET y sobre como las condiciones de aplicación afectan las respuestas, es poco lo que se conoce sobre el proceso de respuesta. Entender mejor este proceso permitiría contribuir a una mejor interpretación y uso de los SET. En este estudio se propone evaluar la hipótesis de que las diferencias individuales en los procesos de respuesta a los SET están relacionadas con variables que median la interpretación cognitiva de las características de la enseñanza, o bien, que explican diferencias individuales en como la enseñanza facilita el aprendizaje. Las principales variables explicativas consideradas fueron: auto-eficacia para aprender, expectativas de resultado, motivación intrínseca, percepción de elección, percepción de valor y utilidad, elección de metas, locus de causalidad, estabilidad y controlabilidad. Se consideraron dos instrumentos básicos: cuestionario de evaluación de la calidad en la enseñanza; y, cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. La muestra de datos de calibración consistió en las respuestas a ambos cuestionarios de 3243 estudiantes, que evaluaron 98 docentes, en 10 cursos, durante 2 semestres consecutivos. La aproximación analítica se basó principalmente en la modelización de las respuestas a los cuestionarios mediante modelos de múltiples facetas de Rasch (MFRM). Los resultados fueron altamente consistentes con la hipótesis de investigación. Palabras claves:

evaluación estudiantil de la enseñanza, teorías socio cognitivas de la motivación, validez, medidas invariantes, modelo de Rasch de multiples facetas.



ABSTRACT

Student Evaluations of Teaching questionnaires (SET) are widely used in universities around the world as a source of feedback on teaching effectiveness. These questionnaires are used as source of data to improve teaching, to make management decisions and administrative, and to study the relationship between teaching and learning. While there have been some studies on the attitudes of students about SETs and how the conditions of application of instruments affect student responses, very little is known about the response process. Better understanding of the process of student response would contribute to a better interpretation and use of SET. In this study we intend to test the hypothesis that individual differences in the response processes of SET are related to variables that mediate cognitive interpretation of the characteristics of teaching or explain individual differences in how teaching facilitates learning in students. The main explanatory variables considered were: self-efficacy to learn, outcome expectations, intrinsic motivation, perception of choice, perception of value and utility, choice of goals, locus of causality, stability and controllability. Two basic instruments were considered: questionnaire for the evaluation of quality in teaching; and, questionnaire on socio-cognitive evaluation of the learning process. The sample of calibration data consisted in the answers to both questionnaires of 3243 students, who evaluated 98 teachers, in 10 courses, during 2 consecutive semesters. The analytical approach was based mainly on the modeling of responses to questionnaires using Rasch's multi-faceted models (MFRM). The results were strongly consistent with the research hypothesis. Keywords:

Student rating of teaching, Sociocognitive theories of motivation, validity, invariante measure, many-facets Rasch model.



Lista de figuras

| | | |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Determinismo recíproco adoptado en la teoría del aprendizaje social de Bandura. | 16 |
| 2.2 | Teoría de la Atribución de Bernard Weiner (reelaborado a partir de Weiner (1985)) | 22 |
| 2.3 | Modelo cognitivo del comportamiento motivado (reelaborado a partir de Deci (1975a)) | 32 |
| 2.4 | Esquema general de la teoría de integración orgánica (reelaborado a partir de R. M. Ryan y Deci (2017)) | 35 |
| 4.1 | Modelo SDT del proceso de respuesta a los SET. Relación tipos de motivación y regulación con el efecto facilitador de la enseñanza y la severidad de los estudiantes. | 70 |
| 4.2 | Modelo SDT del proceso de respuesta a los SET. Relación entre satisfacción de necesidades básicas, efecto facilitador de la enseñanza y severidad de los estudiantes. | 71 |
| 4.3 | Modelo SDT del proceso de respuesta a los SET. Relación del contenido de metas con el efecto facilitador de la enseñanza y la severidad de los estudiantes. | 72 |
| 4.4 | Modelo TAS del proceso de respuesta a los SET. | 73 |
| 5.1 | Número de observaciones de elementos de cada faceta | 88 |
| 5.2 | Gráfica de la faceta docentes y estudiantil. D1, D2 y D3 representan docentes evaluados, mientras que, E1 a E9 representan estudiantes. Los bordes de la gráfica unen los elementos (vértices) conexas entre sí. Entre paréntesis se indica el grado o número de elementos adyacentes o conexiones directas que tiene cada elemento. | 89 |

| | | |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.3 | Conexión entre los elementos de las facetas Docentes y Estudiantes. El grado corresponde al número de elementos al que está conextado cada elemento considerado. | 90 |
| 5.4 | Conexión entre los elementos de las facetas Docentes y Estudiantes. Se muestran las entradas no nulas para la matriz de adyacencia A y sus primeras potencias. Cada punto corresponde a una entrada no nula. Las primeras 3243 filas y columnas corresponden a estudiantes y las últimas 98 filas y columnas corresponden a docentes. | 93 |
| 5.5 | Estabilidad de las estimaciones en función de n para estimaciones de Habilidad, Severidad y Dificultad. | 95 |
| 5.6 | Ejemplos de estabilidad para elementos en cada faceta en función de n | 96 |
| 5.7 | PVE para el modelo en cada paso de eliminación de elementos con falta de ajuste | 97 |
| 5.8 | Comparación de la invarianza de las medidas respecto a los estudiantes según pasos de diagnóstico (1,5 y 10) y métodos (método de Linacre y método alternativo). | 100 |
| 6.1 | Umbrales de Rasch-Aldrich. | 123 |
| 8.1 | Medida de los componentes de severidad para cada sub-escala de evaluación de la calidad en la enseñanza. | 167 |
| 8.2 | Modelos teóricos del proceso de respuesta a los SET y componentes de severidad para las sub-escalas de calidad en la enseñanza. | 170 |
| 8.3 | Agrupamiento de patrones. | 172 |
| 8.4 | Ajuste global de modelos de regresión de la severidad global correspondientes a los modelos teóricos para cada sub-escala. SDT = Modelo SDT, TAS = Modelo TAS, ATR = Modelo Atribucional y AUT = Modelo de Auto-teorías. | 179 |
| 8.5 | Modelos teóricos del proceso de respuesta a los SETs y efectos primarios encontrados. | 193 |
| 9.1 | Variables socio-cognitivas y Proceso de Respuesta. | 202 |
| 9.2 | Relación entre severidad local, global y efecto facilitador. | 203 |

- 9.3 Hipótesis de Validez. Evaluación subjetiva, pero, fidedigna e informativa; refleja diferencias en la calidad del proceso de aprendizaje y el efecto facilitador de la enseñanza. 207



Lista de tablas

| | | |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Relación entre principales variables explicativas, instrumentos de medición y teorías consideradas. | 15 |
| 4.1 | Relación entre Patrones Atribucionales I y II, valores característicos de las dimensiones causales y causas específicas más importantes para cada patrón atribucional. Para cada patrón se muestra entre paréntesis el valor característico para cada dimensión. Además se muestran las causas específicas que podrían ser más importantes para cada patrón. | 68 |
| 4.2 | Modelo de Auto-teorías del proceso de respuesta a los SET | 77 |
| 5.1 | Etapas en el proceso de diagnóstico de falta de ajuste. | 86 |
| 5.2 | Comparación del ajuste global | 98 |
| 6.1 | Ajuste Global entre los datos y el modelo. | 105 |
| 6.2 | Ítems : Medidas y estadísticos de ajuste por ítem | 109 |
| 6.3 | Estadísticos descriptivos para la faceta ítems para cada sub-escala | 110 |
| 6.4 | Confiabilidad de Separación para la faceta Ítems para cada sub-escala | 110 |
| 6.5 | Estadísticos descriptivos para la faceta docentes para cada sub-escala | 112 |
| 6.6 | Confiabilidad de Separación para la faceta Docentes para cada sub-escala | 112 |
| 6.7 | Estadísticos descriptivos para la faceta estudiantes para cada sub-escala | 115 |
| 6.8 | Confiabilidad de Separación para la faceta Estudiantes para cada sub-escala | 115 |
| 6.9 | Análisis de la Invarianza de las Medidas para cada sub-escala . . | 116 |

| | | |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.10 | Análisis de Interacciones (sesgos) entre facetas para cada dimensión considerada. En la Tabla: C (curso), D (docente), E (Estudiante), I (Ítem), * (interacción), Var. Exp (%) (porcentaje de la varianza explicada por la interacción). | 117 |
| 6.11 | Estructura de las escalas de los ítems: Promedio de Medidas . . . | 120 |
| 6.12 | Estructura de las escalas de los ítems: umbrales de Rasch-Andrich | 122 |
| 6.13 | Estructura de las escalas de los ítems: Outfit y Conteo de observaciones no extremas. | 124 |
| 7.1 | Descripción del cuestionario socio-cognitivo de evaluación del proceso de aprendizaje | 133 |
| 7.2 | Ajuste Global entre los datos y el modelo. | 136 |
| 7.3 | Ítems : Medidas y estadísticos de ajuste por ítem | 138 |
| 7.4 | Estadísticos descriptivos para la faceta ítems para cada sub-escala | 139 |
| 7.5 | Confiabilidad de Separación para la faceta Ítems para cada sub-escala. | 139 |
| 7.6 | Estadísticos descriptivos para la faceta Docentes para cada sub-escala. | 141 |
| 7.7 | Confiabilidad de Separación para la faceta Docentes para cada sub-escala. | 141 |
| 7.8 | Estadísticos descriptivos para la faceta Estudiantes para cada sub-escala. | 143 |
| 7.9 | Confiabilidad de Separación para la faceta Estudiantes para cada sub-escala. | 143 |
| 7.10 | Análisis de la Invarianza de las Medidas para cada sub-escala . . . | 144 |
| 7.11 | Análisis de Interacciones (sesgos) entre facetas para cada dimensión considerada. En la Tabla: D (docente), E (Estudiante), I (Ítem), * (interacción), Var. Exp (%) (porcentaje de la varianza explicada por la interacción). | 145 |
| 8.1 | Descripción de los componentes de Severidad en el modelo LLT-MFRM | 155 |
| 8.2 | Modelos teóricos y predicciones teóricas sobre los efectos primarios de las variables explicativas en los modelos estadísticos de la severidad global. CES denota “cambio esperado en severidad”. $CES_{II,I}$ denota el cambio en severidad al pasar del patrón atribucional II al I. | 160 |

| | | |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 8.3 | Ajuste global entre los datos y el modelo LLT-MFRM. Los valores normalizados ($CV(RMSE)$ y $NMBE$) son expresados como porcentaje respecto a la media de los puntajes observados. | 162 |
| 8.4 | Componentes de Severidad. | 165 |
| 8.5 | Componentes de Severidad (continuación). | 166 |
| 8.6 | Confiabilidad de Separación para la faceta Componentes de Severidad para cada sub-escala | 166 |
| 8.7 | Resumen de las medidas de los Componentes de Severidad | 167 |
| 8.8 | Correlaciones (corregidas para atenuación) entre medidas de severidad y medidas de variables socio-cognitivas para cada sub-escala | 173 |
| 8.9 | Valor p (valor p de Holm) para las correlaciones (corregidas para atenuación) entre medidas de severidad y medidas de variables socio-cognitivas para cada sub-escala | 174 |
| 8.10 | Ajuste global entre los datos y modelos regresión correspondientes a cada modelo teórico. La variable respuesta es la severidad global. La selección de las variables explicativas se fundamenta en los modelos teóricos propuestos en el capítulo 4. Los valores normalizados ($NMAE$, $NRMSE$ y $NMBE$) son expresados como porcentaje respecto al rango de la severidad para cada dimensión. F_{Obs} es el valor observado del estadístico para un test F global. p_{Perm} es el valor p de permutación correspondiente al F_{Obs} | 177 |
| 8.11 | Modelo SDT. | 182 |
| 8.12 | Modelo TAS. | 184 |
| 8.13 | Modelo Atribucional. | 189 |
| 8.14 | Modelo de Auto-teorías. | 192 |

Lista de siglas

BNPT Basic Psychological Needs Theory 38

CET Cognitive Evaluation Theory 29

GCT Goal Contents Theory 40

ME Metas extrínsecas 40

MI Metas intrínsecas 40

OIT Organismic Integration Theory 35

SDT Self Determination Theory 28

SET Student Evaluations of Teaching 1

TOTE Test, Operate, Test, Exit 33



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Tabla de contenidos

| | |
|---------------------------------------------------------------------------------------|-------------|
| Lista de figuras | X |
| Lista de tablas | XIII |
| Lista de siglas | XVI |
| 1 Introducción. | 1 |
| 1.1 Cuestionarios de evaluación estudiantil de la enseñanza. | 1 |
| 1.2 Relevancia práctica e institucional de los SET. | 1 |
| 1.3 Problema de investigación. | 2 |
| 1.4 Relevancia del problema de investigación. | 4 |
| 1.5 Organización del la tesis. | 5 |
| 2 Antecedentes y marco teórico. | 6 |
| 2.1 Interpretación de los SET. | 6 |
| 2.2 Dimensiones generales de efectividad o calidad en la enseñanza. | 7 |
| 2.3 Investigaciones sobre SET. | 9 |
| 2.3.1 Validez factorial de los SET. | 10 |
| 2.3.2 Proceso de respuesta y validez de los SET. | 11 |
| 2.4 Teorías cognitivas de la motivación y proceso de respuesta a los SET. | 13 |
| 2.4.1 Teorías consideradas, principales variables e instrumentos de medición. | 13 |
| 2.4.2 Teoría del Aprendizaje Social de Albert Bandura. | 16 |
| 2.4.3 Teoría de la Atribución de Bernard Weiner. | 19 |
| 2.4.4 Teoría de la autodeterminación. | 28 |
| 2.4.5 Modelo de Dweck y Leggett. | 41 |
| 2.5 Modelos de Rasch y estudio del proceso de respuesta a los SET. | 43 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3 Metodología. | 48 |
| 3.1 Muestra. | 48 |
| 3.1.1 Funcionamiento del Sistema de Evaluación y datos recabados a partir del año 2011. | 48 |
| 3.1.2 Criterios de selección de la muestra de datos. | 50 |
| 3.1.3 Muestra de datos de calibración. | 51 |
| 3.2 Instrumentos. | 52 |
| 3.3 Procedimientos. | 53 |
| 3.3.1 Pre-procesamiento de los datos y anonimización. | 53 |
| 3.3.2 Análisis de Datos. | 54 |
| 3.3.3 Consideraciones Éticas. | 62 |
| | |
| 4 Hacia una teoría socio-cognitiva del proceso de respuesta a cuestionarios de evaluación estudiantil de la calidad de la enseñanza. | 64 |
| 4.1 Presentación. | 64 |
| 4.2 Modelo atribucional para el proceso de respuesta a los SET. | 65 |
| 4.3 Modelo SDT del proceso de respuesta a los SET. | 68 |
| 4.4 Modelo TAS del proceso de respuesta a los SET. | 72 |
| 4.5 Modelo de Auto-teorías del proceso de respuesta a los SET. | 74 |
| 4.6 Discusión. | 78 |
| | |
| 5 Diagnóstico de modelos de múltiples facetas de Rasch en estudios observacionales. | 81 |
| 5.1 Presentación. | 81 |
| 5.2 El Modelo de medición de múltiples facetas de Rasch | 82 |
| 5.3 Muestra de datos de calibración. | 83 |
| 5.4 Diagnóstico de ajuste entre datos y el modelo. | 84 |
| 5.5 Método de diagnóstico alternativo propuesto. | 84 |
| 5.6 Resultados. | 86 |
| 5.6.1 Caracterización de la muestra de datos de calibración. | 86 |
| 5.6.2 Evaluación empírica del método de diagnóstico. | 94 |
| 5.7 Sumario y Conclusiones. | 100 |
| | |
| 6 Validez de un cuestionario de evaluación estudiantil de la calidad en la enseñanza: análisis de múltiples facetas de Rasch. | 102 |
| 6.1 Presentación. | 102 |

| | | |
|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 6.2 | Metodología. | 103 |
| 6.2.1 | Muestra de datos de calibración. | 104 |
| 6.2.2 | Instrumentos. | 104 |
| 6.2.3 | Aproximación Analítica. | 104 |
| 6.3 | Resultados. | 104 |
| 6.3.1 | Ajuste global entre datos y el modelo. | 104 |
| 6.3.2 | Análisis de las Facetas de Docentes, Estudiantes e Ítems. | 106 |
| 6.3.3 | Invarianza. | 116 |
| 6.3.4 | Interacciones (sesgos) entre los elementos de las facetas. | 117 |
| 6.3.5 | Estructura de las escalas de los ítems y uso de las categorías. | 118 |
| 6.4 | Discusión. | 126 |
| 7 | Validez de un cuestionario de evaluación socio-cognitiva del proceso de aprendizaje: análisis de múltiples facetas de Rasch. | 131 |
| 7.1 | Presentación. | 131 |
| 7.2 | Metodología. | 132 |
| 7.2.1 | Muestra de datos de calibración. | 132 |
| 7.2.2 | Instrumentos. | 132 |
| 7.2.3 | Aproximación Analítica. | 133 |
| 7.3 | Resultados. | 135 |
| 7.3.1 | Ajuste global entre datos y el modelo. | 135 |
| 7.3.2 | Análisis de las facetas de Docentes, Estudiantes e Ítems. | 137 |
| 7.3.3 | Invarianza. | 144 |
| 7.3.4 | Interacciones (sesgos) entre los elementos de las facetas. | 145 |
| 7.4 | Discusión. | 147 |
| 8 | Modelado del proceso de respuesta a cuestionarios de evaluación estudiantil de la calidad en la enseñanza aplicando teorías socio-cognitivas de la motivación. | 151 |
| 8.1 | Presentación | 151 |
| 8.2 | Metodología | 153 |
| 8.2.1 | Muestra de datos de calibración. | 153 |
| 8.2.2 | Instrumentos | 153 |
| 8.2.3 | Aproximación Analítica | 154 |
| 8.3 | Resultados | 161 |

| | | |
|----------|----------------------------------------------------------------------------------------|------------|
| 8.3.1 | Ajuste global entre datos y el modelo LLT-MFRM. | 161 |
| 8.3.2 | Análisis de la Faceta Componentes de Severidad | 163 |
| 8.3.3 | Evaluación de la hipótesis de investigación: análisis correlacional | 173 |
| 8.3.4 | Evaluación de la hipótesis de investigación: modelado de la severidad global | 175 |
| 8.4 | Discusión | 194 |
| 9 | Consideraciones finales. | 198 |
| 9.1 | Hipótesis de investigación y evidencias encontradas. | 198 |
| 9.2 | Validez de constructo del cuestionario SET. | 203 |
| 9.3 | Validez de constructo del cuestionario SC. | 205 |
| 9.4 | Modelos teóricos del proceso de respuesta e Hipótesis de Validez. | 206 |
| 9.5 | Aplicaciones prácticas. | 207 |
| | Referencias | 208 |
| | Glosario | 224 |
| | Apéndices | 225 |
| | Apéndice 1 Consentimiento Informado | 226 |
| | Apéndice 2 Código empleado en la anonimización de datos. | 228 |

Capítulo 1

Introducción.

1.1. Cuestionarios de evaluación estudiantil de la enseñanza.

Los cuestionarios de evaluación estudiantil de la enseñanza (“Student Evaluation of Teaching” o SET) son instrumentos de evaluación de la enseñanza utilizados en universidades y otras instituciones de educación superior. Varios ejemplos importantes de SET son escalas organizadas en subescalas relacionadas con diferentes dimensiones de la enseñanza (Marsh, 1987).

Los ítems empleados en estas escalas incluyen ítems cerrados. Estos ítems presentan categorías de respuesta ordinales. Por ejemplo, algunas escalas de ítems utilizadas indican diferentes niveles de acuerdo respecto a una afirmación. Estos instrumentos pueden contener también preguntas abiertas que pueden solicitar, por ejemplo, que el estudiante describa fortalezas y debilidades de la enseñanza, o señale aspectos a mejorar (Arreola, 2006; Braskamp y Ory, 1994; Marsh, 1987; Seldin, 2006; Theall, Abrami, y Mets, 2001).

1.2. Relevancia práctica e institucional de los SET.

Los SET son ampliamente usados en universidades de todo el mundo como una fuente de retroalimentación acerca de la efectividad o calidad en la enseñanza. Estos cuestionarios son empleados como fuente de datos para mejorar la enseñanza, para tomar decisiones de gestión y administrativas, y para

estudiar la relación entre enseñanza y aprendizaje (Brown, 2011; Madichie, 2011; Marsh, 1987; Seldin, 1989; Theall y cols., 2001; R. E. Wright, 2006).

En el caso de la Universidad de la República, en el año 2012, el Consejo Directivo Central aprobó lo siguiente: “Asimismo, cada servicio deberá diseñar correctamente e instrumentar mecanismos representativos de evaluación estudiantil de la actividad docente de forma que éstos puedan ser tenidos en cuenta al momento de la evaluación docente.” (UDELAR, 2012).

En la Facultad de Medicina de la Universidad de la República estos instrumentos son la fuente principal de datos para el Sistema de Evaluación de la Enseñanza y Aprendizaje. Este sistema se está desarrollando en consonancia con políticas nacionales (URUGUAY, 2008) y de la Universidad de la República (UDELAR, 2012).

Los propósitos básicos de este sistema de evaluación son: (a) la mejora de la calidad de la enseñanza y aprendizaje a través de retroalimentaciones a docentes y estudiantes acerca de ambos procesos; y (b) servir como un sistema de información que aporte datos necesarios para entender la relación entre enseñanza y aprendizaje, en el contexto institucional en el que se desarrollan ambos procesos.

En consecuencia, los SET cumplen en Facultad de Medicina un papel clave en el sistema de evaluación institucional de la enseñanza y proceso de aprendizaje. De esta forma, los SET desempeñan un papel importante en el diagnóstico y promoción de la calidad de los procesos de enseñanza y aprendizaje. Por otra parte, la implantación de este sistema es un aspecto importante en los procesos de evaluación institucional y de acreditación de carreras, en especial porque permite un diagnóstico de la calidad de los procesos de enseñanza y aprendizaje. Por lo tanto, los SET pueden tener en estos procesos de evaluación y acreditación institucional también un papel destacado.

1.3. Problema de investigación.

En esta investigación se propone evaluar la hipótesis de que las diferencias individuales en los procesos de respuesta a los SET obedecen a dos posibles efectos: por un lado, un efecto vinculado a diferencias relacionadas con características del individuo que median la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza; por otro lado, un efecto concerniente a diferencias individuales relacionadas con formas distintas de aprender y con

un efecto diferencial de la enseñanza sobre el proceso de aprendizaje del estudiante.

Las principales variables explicativas consideradas en la evaluación de la hipótesis formulada son las siguientes: tipos de atribuciones, auto-eficacia, expectativa de resultado, motivación intrínseca, percepción de elección, valor/utilidad y elección de metas. Estas variables son centrales en teorías cognitivas de la motivación, y están involucradas en el proceso de aprendizaje.

Esta hipótesis central se justifica a partir de evidencias previas y consideraciones teóricas. En primer lugar, varios resultados de trabajos previos son consistentes con esta hipótesis (R. Ames y Lau, 1979; Grimes, Millea, y Woodruff, 2004). En segundo lugar, para la formulación de esta hipótesis considero también argumentos teóricos basados en teorías cognitivas de la motivación (Bandura, 1977; Schunk, Meece, y Pintrich, 2014). Estos argumentos intentan mostrar que la hipótesis propuesta puede deducirse de la aplicación de teorías cognitivas de la motivación a la evaluación estudiantil de la enseñanza.

En la presente investigación planteo dos preguntas claves en relación con la evaluación de la hipótesis formulada. Mi pregunta de investigación principal es: ¿estas variables permiten explicar diferencias individuales en los procesos de respuesta a los SET? Por ejemplo: ¿estudiantes que realizan “atribuciones internas” se diferencian en el proceso de respuesta de los estudiantes que realizan “atribuciones externas”?

La segunda pregunta de investigación, relacionada con la hipótesis propuesta, se origina a partir de las siguientes observaciones. En primer lugar, podrían existir diferencias en cómo las variables explicativas consideradas median la interpretación de señales utilizadas para evaluar diferentes dimensiones de la enseñanza. En segundo lugar, podrían existir diferencias en cómo las variables explicativas consideradas median la forma en que diferentes dimensiones de la enseñanza afectan el aprendizaje de los estudiantes.

Considerando estas dos observaciones, se plantea también esta segunda pregunta de investigación: ¿Existen diferencias en cómo estas variables median el proceso de respuesta con relación a la evaluación de diferentes dimensiones de la enseñanza? Por ejemplo: ¿Existen diferencias en cómo media la motivación intrínseca la interpretación cognitiva de aspectos de la enseñanza cuando se considera la dimensión “valor de la enseñanza” que cuando se considera la dimensión “claridad y organización”.

1.4. Relevancia del problema de investigación.

Las respuestas a las preguntas planteadas en la investigación aquí presentada tienen varias implicaciones. En primer lugar, aportan un abordaje sustantivo a las evidencias de validez de los SET que es señalada como una carencia con relación a los estudios de validez (Theall y cols., 2001). Este abordaje sobre la validez, a su vez, se espera que permita mejorar las prácticas de aplicación de estos instrumentos y que contribuyan a una mejora de su utilidad.

En segundo lugar, responder a las preguntas planteadas puede también contribuir a una comprensión más profunda acerca de la interacción entre enseñanza y aprendizaje. Esta comprensión más profunda, por su parte, contribuiría a establecer una conexión entre evaluaciones de la enseñanza enfocadas en las características de los cursos y docentes, con un enfoque orientado hacia el estudiante y su proceso de aprendizaje.

En tercer lugar, las respuestas a estas interrogantes, pueden aportar insumos para construir una noción de efectividad en la enseñanza que reconoce y operacionaliza que el efecto de la enseñanza sobre el aprendizaje podría variar dependiendo de las características individuales de los sujetos.

Además de las implicaciones antes señaladas, el presente estudio tiene una gran significación para la Facultad de Medicina de la Universidad de la República. Hasta el momento, las evidencias de validez de los instrumentos utilizados se basan en consideraciones sobre validez de contenido, que si bien son importantes, no son evidencias suficientes de la validez de estos instrumentos.

Otro aspecto por el cual esta investigación es de gran relevancia para la Facultad de Medicina, está relacionada con las características específicas del contexto y condiciones en que se enseña y evalúa la enseñanza en esta institución. Existen importantes diferencias con respecto a otras Facultades y Escuelas de Medicina en Universidades de todo el mundo. Estas diferencias implican una fuerte necesidad de producir nuevos conocimientos teóricos y desarrollos metodológicos. Los resultados de la presente investigación son de suma importancia para lograr satisfacer esa necesidad institucional.

En suma, los resultados obtenidos por esta investigación se espera que aporten evidencias sustantivas de validez de los instrumentos empleados en el Sistema de Evaluación de la Enseñanza y Aprendizaje de la Facultad de Medicina. También se espera que los resultados obtenidos signifiquen aportes de nuevos conocimientos teóricos, desarrollos metodológicos, y procedimientos de

evaluación válidos, claves para el contexto en el que se desarrolla la enseñanza y evaluación educativa en esta institución.

Así, todos estos aportes serán claves para dar respuesta a las necesidades institucionales vinculadas con la mejora de la calidad de la enseñanza y el aprendizaje. También, se espera que estos aportes tengan un impacto substancial en la mejora de la forma en que se lleva a cabo la evaluación educativa en Facultad de Medicina. Más aún, debido a la riqueza y complejidad de esta institución, probablemente, un abordaje exitoso del problema de investigación propuesto y el impacto de sus resultados, tendrán a su vez, un impacto positivo en el desarrollo de la investigación y evaluación educativa en toda la Universidad de la República.

Estos resultados esperados son claves desde el punto de vista aplicado. No obstante, la resolución del problema de investigación planteado, puede representar también una contribución científica relevante. Esta contribución está relacionada con problemas de investigaciones que no han podido ser resueltos adecuadamente en investigaciones previas realizadas hasta el momento.

1.5. Organización del la tesis.

En el capítulo 2 se describe el marco teórico de esta tesis y su relación con la hipótesis de investigación. También, se sintetizan antecedentes relevantes para el problema abordado. En el capítulo 3 se describe el marco metodológico general adoptado. Los capítulos 4, 5, 6, 7 y 8 presentan resultados originales aportados en esta tesis. En el capítulo 4 a partir de cuatro perspectivas teóricas consideradas y de la hipótesis central de esta investigación se proponen varios modelos teóricos sobre el proceso de respuesta a los cuestionarios SET. En el capítulo 5 se aborda un problema metodológico que surgió durante la presente investigación. En el capítulo 6 y 7 se aportan evidencia de validez del cuestionario de evaluación estudiantil de la calidad de la enseñanza (cuestionario SET) y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje (cuestionario SC). En estos dos capítulos, adicionalmente, se realiza una primera evaluación de la hipótesis central de esta investigación. En el capítulo 8 se lleva a cabo una evaluación más directa de la hipótesis central de esta tesis, así como también, se aportan nuevas evidencias de la validez de constructo del cuestionario SET y SC. Finalmente, en el capítulo 9 se discuten los resultados de esta tesis y su implicaciones teóricas y prácticas.

Capítulo 2

Antecedentes y marco teórico.

2.1. Interpretación de los SET.

La interpretación de los datos recabados con SET depende de la concepción de enseñanza que se considere (Arreola, 2006). El autor de esta tesis adhiere a una conceptualización de la enseñanza que considera a esta un proceso de interacción con el estudiante, diseñado específicamente para facilitar su aprendizaje.

En esta conceptualización de la enseñanza, no se asume que la enseñanza es una causa suficiente de aprendizaje del estudiante. Así, en esta perspectiva el estudiante tiene un papel fundamental en su proceso de aprendizaje.

De acuerdo con esta concepción de la enseñanza, la efectividad o calidad en la enseñanza está relacionada con características de la enseñanza que facilitan o promueven el aprendizaje del estudiante (J. Andrews, 2004). En ese sentido la enseñanza puede ser más o menos efectiva dependiendo de la presencia de ciertas características y de cuál es el efecto de estas sobre el aprendizaje.

Consistente con esta concepción de la enseñanza, se supone fundamental que la medición y evaluación de efectividad en la enseñanza se realice considerando que el proceso de enseñanza y el proceso de aprendizaje presentan una interacción compleja influenciada por variables contextuales (J. Andrews, 2004; Bandura, 1977).

Otro aspecto clave en la interpretación de los SET es el papel atribuido a los estudiantes en la evaluación de la enseñanza (Arreola, 2006). La evaluación puede definirse como la interpretación de una medida que permite determinar en qué grado cierto valor obtenido representa una condición deseada. En esta

investigación se adhiere a una conceptualización de los SET como una medida del impacto de la enseñanza en los estudiantes. Así, este impacto se mide considerando el grado de acuerdo de los estudiantes con respecto a una serie de afirmaciones vinculadas a contenidos relacionados con la enseñanza que pueden ser observados por los estudiantes.

Con esta perspectiva, la interpretación de los SET requiere procedimientos de evaluación cuya validez debe ser previamente investigada. Dentro de este enfoque los patrones de respuesta de los estudiantes a los SET aportan datos que pueden ser analizados para extraer información sobre la efectividad en la enseñanza y en general sobre el impacto de la enseñanza en los estudiantes. Así, la evaluación de la enseñanza, con el propósito de hacer un diagnóstico y promover mejoras, no es simplemente una mera síntesis o descripción de las respuestas de los estudiantes a los cuestionarios. Sino, por el contrario, es una evaluación realizada posteriormente cuando se aplica un procedimiento válido de evaluación que permite diagnosticar en qué medida los patrones de respuesta obtenidos revelan fortalezas o debilidades de la enseñanza.

Dentro del enfoque abordado, además del uso de SET, las formas de puntajes estudiantiles de autoevaluación del proceso de aprendizaje son también instrumentos útiles para: extraer información sobre la efectividad en la enseñanza; sobre el impacto de la enseñanza sobre el proceso de aprendizaje de los estudiantes; y para analizar la validez de medidas y procedimientos de evaluación basados en SET.

2.2. Dimensiones generales de efectividad o calidad en la enseñanza.

Arreola (2006) distingue cinco dimensiones generales de habilidad requeridas para llevar a cabo una enseñanza efectiva. La definición de estas dimensiones parte del análisis de aspectos de la actividad de enseñanza que están implicadas en lograr que el docente interactúe con el estudiante para: proveer la oportunidad de aprender; crear condiciones que facilitan el aprendizaje; y, utilizar técnicas y métodos que crean una alta probabilidad de que el aprendizaje ocurra.

En primer lugar, define la Experticia en el Contenido como “el cuerpo de habilidades, competencias, y conocimientos específicos, en el área o tópico

específico a ser enseñado, en el cual el docente ha recibido educación avanzada, entrenamiento, y/o experiencia” (p. 20).

En segundo lugar, considera las Habilidades de Diseño o “aquellas habilidades técnicas en el diseño y secuenciación de experiencias que, cuando son realizadas por un estudiante calificado, resultan en una alta probabilidad que los resultados de aprendizaje específico sean alcanzados por el estudiante” (p. 20).

En tercer lugar, define las Habilidades de Impartir Instrucción como las “habilidades de interacción humana y características que facilitan el aprendizaje, incluyendo la habilidad para motivar a los estudiantes, generar entusiasmo, y comunicación efectiva usando varias formas de tecnologías de transmisión de información” (p. 20).

En cuarto lugar, considera las Habilidades de Evaluación o aquellas “habilidades en el desarrollo, procedimientos, y estrategias para la evaluación del aprendizaje de los estudiantes y provisión de una retroalimentación significativa” (p. 24).

Por último, define las habilidades de gestión como aquellas “habilidades en configurar, mantener, y administrar recursos e instalaciones requeridas para proveer un ambiente apropiado para la enseñanza y aprendizaje” (p. 24).

Las dimensiones generales definidas por [Arreola \(2006\)](#) resultan útiles como una primera aproximación. No obstante, puede ser útil considerar también algunas visiones sobre la enseñanza y ambientes de aprendizaje, que hacen más explícitos varios aspectos generales vinculados con la efectividad o calidad en la enseñanza.

Por ejemplo, [Braskamp y Ory \(1994\)](#) señalan que los métodos de enseñanza están inmersos en el contenido de una disciplina, y por tanto, deberían ser vistos como dos aspectos integrados e interrelacionados. Adicionalmente, [Braskamp y Ory \(1994\)](#) también destacan la importancia de habilidades vinculadas a tareas de asesoramiento, supervisión, orientación y tutoría, que pueden distinguirse en alguna medida de habilidades circunscritas a la interacción en el aula. Además, dentro de una visión más inclusiva de la actividad de enseñanza, consideran que una dimensión de la efectividad en la enseñanza está vinculada con la habilidad de desarrollarse como docente. En relación a este aspecto específico, [Braskamp y Ory \(1994\)](#) mencionan como actividades vinculadas a esta habilidad general la evaluación de la enseñanza de docentes, el llevar a cabo investigación sobre la enseñanza y el asistir a actividades de desarrollo

profesional.

Por su parte, [Theall y cols. \(2001\)](#) destacan la importancia de considerar principios centrados en el aprendizaje en la definición de aspectos que definen qué es una enseñanza efectiva. Desde esta perspectiva, por ejemplo, se hacen explícitas ciertas habilidades requeridas, como por ejemplo, habilidades vinculadas a promover el compromiso de los estudiantes en un aprendizaje auto-regulado; o bien, la habilidad para facilitar un aprendizaje cooperativo.

Otra aproximación que puede aportar una visión general acerca de diferentes dimensiones generales de la efectividad en la enseñanza esta relacionada con la definición de competencias para la educación desarrollada en el marco del Proyecto Tunning América Latina ([Coronado, 2009](#)). No obstante, muchas de las competencias definidas tienen una relación lógica más o menos directa con las dimensiones generales definidas por [Arreola \(2006\)](#). En forma similar, algunas de las competencias definidas explicitan también algunos de los aspectos señalados por [Braskamp y Ory \(1994\)](#) y por [Theall y cols. \(2001\)](#).

2.3. Investigaciones sobre SET.

Existen varios miles de trabajos relacionados con SET que se focalizan en el estudio de las respuestas a los ítems cerrados. Estos estudios en general consideran como variable básica el promedio de los puntajes para cada ítem en cada clase o grupo de estudiantes, que evalúan un curso o docente.

Algunas de las conclusiones principales de estas investigaciones expresan que los SET son: (a) multidimensionales, esto es, permiten medir un constructo formado por varias dimensiones; (b) son confiables y estables; (c) relativamente validos en relación con otras medidas de efectividad de la enseñanza; (d) relativamente no afectados por variables consideradas como posibles sesgos; y (e) son útiles para mejorar el desempeño docente, en especial cuando se acompañan de asesoramiento o estrategias de mejora ([Arreola, 2006](#); [Marsh, 1987](#); [Marsh y Hocevar, 1991](#); [Theall y cols., 2001](#)).

En relación con los puntos anteriores es importante reconocer que esos resultados refieren a puntajes promedio para cada ítem por cada grupo de estudiantes a un docente o curso. El punto (a) resulta de la aplicación de técnicas de análisis factorial exploratorio a los puntajes promedios de ítems y no a los puntajes individuales que asigna cada estudiante ([Marsh y Hocevar, 1991](#)).

En forma similar, en el punto (b) refiere a la confiabilidad de los puntajes promedio. Por ejemplo, Marsh (1987) observa que un valor típico de la confiabilidad para el cuestionario de evaluación estudiantil SEEQ es alrededor de 0.95 para cada factor, para respuestas promedio obtenidas a partir de 50 estudiantes, mientras que para 1 estudiante es solo de 0.23.

2.3.1. Validez factorial de los SET.

Marsh, Tourón, y Wheeler (1985) aportaron evidencias sobre la validez factorial de los SET en una población de habla hispana. En esta investigación consideraron dos instrumentos “Endeavor Instrument” (Frey, Leonard, y Beatty, 1975) y “Students’ Evaluation of Educational Quality” o SEEQ (Marsh, 1987). Marsh y cols. (1985) encontraron que ambos instrumentos, diseñados originalmente para poblaciones anglosajonas, permiten medir las dimensiones originalmente definidas, en un contexto e idioma diferente.

Se encontraron también evidencias que ambos instrumentos permiten medir varias dimensiones esencialmente iguales. El análisis conjunto de los ítems de ambos instrumentos permitió definir en total 10 dimensiones. Una de las dimensiones encontradas está relacionada con Habilidades de Evaluación. Las demás dimensiones halladas por Marsh y cols. (1985) están relacionadas con las Habilidades para Impartir Instrucción.

Marsh y Hocevar (1991) aportaron también evidencias de que las dimensiones obtenidas pueden ser generalizadas entre disciplinas, niveles de los docentes y niveles de los cursos. Mientras que, otros trabajos (Marsh, 1981; Watkins, 1992; Watkins y Akande, 1992; Watkins, Marsh, y Young, 1987) han mostrado la generalidad de las dimensiones encontradas en diferentes contextos universitarios en países como por ejemplo: Australia, Nueva Zelandia, Nigeria y China.

Estas investigaciones muestran que en forma consistente se encuentran en conjunto 10 dimensiones: (1) “Valor/Logro/Aprendizaje”; (2) “Entusiasmo del docente”; (3) “Claridad”; (4) “Organización/Planificación”; (5) “Interacción con el grupo/Discusión”; (6) “Relación personal”; (7) “Profundidad”; (8) “Evaluación”; (9) “Tareas asignadas”; y (10) “Carga de trabajo/Dificultad”.

Los resultados de estudios de análisis factorial de puntajes promedio por grupo de estudiantes sugieren que estos grupos pueden distinguir las diferentes dimensiones de enseñanza consideradas (Marsh, 1987). No obstante, esta

técnica no permite estudiar diferencias individuales en la capacidad para distinguir distintas dimensiones de la enseñanza y su efecto sobre el proceso de respuesta.

Así, los factores encontrados con el análisis factorial de promedios de clases pueden reflejar en parte la covariación observada entre las diferentes dimensiones de la enseñanza y una mezcla desconocida de teorías implícitas de los estudiantes acerca de cómo dichas dimensiones covarían (Marsh, 1987).

Cuando un estudiante está evaluando a un docente con el que está poco familiarizado, y dispone de poca información, el estudiante puede recurrir en mayor grado a un esquema relacional o estructura de conocimiento sobre la relación entre diferentes dimensiones de la enseñanza (Renaud y Murray, 2005).

En conclusión, estas investigaciones sobre validez factorial muestran la generalidad de los factores encontrados para promedios de puntajes por grupo, para diferentes idiomas y contextos universitarios. También, reconocen la importancia de aspectos tales como las teorías implícitas, esquemas relacionales o estructuras de conocimiento. Sin embargo, el análisis factorial de puntajes promedio por grupo no permite estudiar variaciones en las teorías implícitas de los estudiantes, ni cómo estas afectan su proceso de respuesta. Por lo que estas investigaciones, si bien reconocen la importancia de estos aspectos, no lograron abordarlos adecuadamente.

2.3.2. Proceso de respuesta y validez de los SET.

A pesar de la importancia práctica que tienen los SET, y de la gran cantidad de trabajos de investigaciones sobre estos instrumentos, existen algunos vacíos importantes de conocimiento con relación a su validez. Si bien se han realizado algunos estudios sobre las actitudes de los estudiantes acerca de los SET (Dwinell y Higbee, 1993; Marlin Jr, 1987) y sobre cómo las condiciones de aplicación de los instrumentos afectan las respuestas de los estudiantes (Centra, 1976; Feldman, 1979), es muy poco lo que se conoce sobre el proceso de respuesta.

Entender mejor el proceso de respuesta de los estudiantes aportaría evidencias adicionales de validez que permitirían contribuir a una mejor interpretación y uso de los SET (Theall y cols., 2001). No obstante, hasta el presente en las investigaciones sobre validez de los SET predomina una aproximación de validez de constructo, basada en el análisis de correlaciones de los puntajes ob-

tenidos con otras variables (Marsh, 1987). En estas investigaciones de validez de constructo no se considera el proceso de respuesta.

Por el contrario, otra aproximación al estudio de la validez se funda en el análisis de la representación de constructo, que concierne los procesos, estrategias y estructuras de conocimiento que están involucradas en la respuesta a los ítems de las escalas consideradas (Embretson, 1994). Esta aproximación está comprendida dentro de un enfoque sustantivo a la validez de constructo. Aspectos sustantivos de la validez de constructo involucran evidencias que soportan el análisis teórico y empírico de los procesos, estrategias y conocimientos que están involucrados en el proceso de respuesta a los ítems de una escala (Messick, 1995).

Un aspecto importante para aproximarse al problema de desentrañar el proceso de respuesta a los SET está relacionado con encontrar variables que puedan mediar en el proceso. R. Ames y Lau (1979) en un estudio que aplicó la Teoría de la Atribución (Weiner, 1974) a la validez de los SET aportan algunas evidencias valiosas para el estudio del proceso de respuesta.

R. Ames y Lau (1979) consideraron como variable explicativa las atribuciones que realizan los estudiantes. Una atribución es una explicación causal que realiza un sujeto de un suceso que le ocurre en su vida. R. Ames y Lau (1979) encontraron que atribuciones internas del desempeño están significativamente relacionadas con evaluaciones positivas, mientras que atribuciones externas del desempeño están significativamente relacionadas con evaluaciones negativas.

R. Ames y Lau (1979) aportaron evidencias de que las atribuciones causales están implicadas en el proceso de respuesta aplicando modelos de regresión múltiple. En estos modelos consideraron como variable explicativa los promedios de grupos de estudiantes para cada ítem del instrumento y para cada sub-escala. Si bien esto aporta evidencias acerca de la importancia de las atribuciones causales que hacen los estudiantes acerca de su aprendizaje y logro, la metodología empleada no permite evaluar a un nivel individual como se relacionan las atribuciones con el proceso de respuesta de los estudiantes. No obstante, los resultados obtenidos por R. Ames y Lau (1979) sugieren que el proceso de evaluación está mediado por una interpretación cognitiva relacionada con atribuciones causales que hacen los estudiantes.

Otro antecedente que se aproxima a estudiar variables que pueden mediar el proceso de respuesta es el trabajo de Grimes y cols. (2004). Estudiaron la relación entre el “locus de control” de los estudiantes y los puntajes en los SET.

El locus de control está relacionado con el grado en que un individuo percibe que un resultado personal depende de sus propios comportamientos o atributos versus el grado en el cual siente que el resultado personal es controlada por factores externos a sí mismo y que puede ocurrir independientemente de sus propias acciones (Rotter, 1966, 1990).

Grimes y cols. (2004) encontraron que estudiantes con un locus de control más internamente orientado tienen una mayor probabilidad de asignar puntajes por encima del promedio, mientras que estudiantes con un locus de control más externamente orientados tienen una mayor probabilidad de realizar evaluaciones promedio o por debajo del promedio. Si bien estos autores consideraron un modelo probit ordinal de las respuestas de cada estudiante, solo consideraron una evaluación global realizada con un único ítem por aspecto evaluado: desempeño del docente, aprendizaje del estudiante y evaluación global del curso.

2.4. Teorías cognitivas de la motivación y proceso de respuesta a los SET.

Esta sección está destinada a dar una visión panorámica de las teorías cognitivas de la motivación consideradas en esta investigación y su relación con la hipótesis propuesta. En particular, se revisan la Teoría del Aprendizaje Social de Bandura (Bandura y Walters, 1977), la Teoría de la Atribución (Weiner, 1974, 1985; Weiner, Russell, y Lerman, 1979), la Teoría de la Auto-determinación (Deci y Ryan, 2008; Deci, Ryan, y Williams, 1996; Deci, Vallerand, Pelletier, y Ryan, 1991; R. M. Ryan y Deci, 2000b), las Teorías de Orientación a Metas y Auto-teorías (C. Ames, 1992; C. Ames y Archer, 1988; Dweck, 1986, 2000; Dweck y Leggett, 1988; Nicholls, 1984; Pintrich, 2000; Schunk y cols., 2014).

2.4.1. Teorías consideradas, principales variables e instrumentos de medición.

Además de las atribuciones causales, mencionada en la sección anterior como una variable potencialmente importante para explicar el proceso de respuesta a los SET, se consideran otras variables explicativas. Estas otras varia-

bles, consideradas variables explicativas principales, se piensa que pueden ser relevantes para entender los procesos de respuesta a los SET. Adicionalmente, pueden ser importantes para desarrollar un concepto de efectividad en la enseñanza que considere las diferencias individuales acerca de cómo la enseñanza facilita el aprendizaje.

Las principales variables explicativas en esta investigación, al igual que el tipo de atribuciones, están relacionados con teorías cognitivas de la motivación (Schunk y cols., 2014). Estas teorías consideran cómo las creencias, actitudes y otras cogniciones, se relacionan con procesos motivacionales. Asimismo, las teorías cognitivas de la motivación no solo intentan explicar los procesos motivacionales, sino también, aportar explicaciones acerca de cómo estos procesos afectan el aprendizaje.

Por tanto, estas teorías ofrecen un marco conceptual y teórico valioso para entender cómo las diferencias individuales, en relación a las variables consideradas, podrían explicar diferencias de interpretación cognitiva de características de la enseñanza, y en consecuencia, mediar procesos de respuesta a los SET. Adicionalmente, también estas teorías podrían ayudar a explicar cómo las diferencias individuales relacionadas con estas variables podrían explicar las diferencias en cómo la enseñanza facilita el aprendizaje en los estudiantes.

En la Tabla 2.1 se muestra la relación entre las principales variables explicativas consideradas en esta investigación, los instrumentos de medición aplicados, y las teorías cognitivas de la motivación consideradas. Cada teoría hace énfasis en algunos aspectos, y toma como variables centrales algunas de las variables consideradas.

Tabla 2.1: Relación entre principales variables explicativas, instrumentos de medición y teorías consideradas.

| Principales variables explicativas | Instrumentos de medición aplicados | Principales teorías consideradas |
|----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Tipo de atribuciones | Escala derivada de la aplicada por R. Ames y Lau (1979) | Teoría de la Atribución Graham (1991) Schunk y cols. (2014) Weiner (1974) |
| 2. Expectativa de Resultados | Escala referida a los resultados esperados en términos de aprendizaje y calificaciones. | Teoría del Aprendizaje Social de Bandura; Bandura (1977) Bandura y Walters (1977) Bandura (2001) Gredler (2009) |
| 3. Auto-eficacia, (o Percepción de Competencia en la Teoría de la Auto-determinación) | Escalas derivada de la sub-escala de Auto-Eficacia para Aprendizaje y Desempeño Pintrich y cols. (1991) y de la sub-escala de Competencia Percibida para Aprender G. Williams, Ryan, y Deci (2011a, 2011b); G. C. Williams y Deci (1996) | Teoría del Aprendizaje Social de Bandura y Teoría de la Auto-determinación Deci y cols. (1991) R. M. Ryan y Deci (2000b) Deci y cols. (1996) Deci y Ryan (2008) R. M. Ryan y Deci (2000b) |
| 4. Motivación Intrínseca | Escala derivada de la sub-escala Interes/Disfrute R. Ryan (2011); R. M. Ryan (1982) | Teoría de la Auto-determinación |
| 5. Percepción de Elección | Escala derivada de la sub-escala Percepción de Elección R. Ryan (2011); R. M. Ryan (1982) | Teoría de la Auto-determinación |
| 6. Valor/Utilidad | Escala derivada de la sub-escala Valor/Utilidad R. Ryan (2011); R. M. Ryan (1982) | Teoría de la Auto-determinación |
| 7. Elección de Metas | Escala derivada de Cuestionario de Elección de Metas Dweck (2000) | Teorías de Orientación a Metas y Auto-Teorías C. Ames y Archer (1988) C. Ames (1992) Dweck (1986) Dweck y Leggett (1988) Dweck (2000) Nicholls (1984) Pintrich (2000) Schunk y cols. (2014) |

2.4.2. Teoría del Aprendizaje Social de Albert Bandura.

De acuerdo a la teoría del Aprendizaje Social de Bandura (Bandura, 1977, 1978b; Bandura y Walters, 1977) el aprendizaje es el resultado de tres componentes básicos que interactúan: los factores personales, los factores conductuales y los factores ambientales (véase Figura 2.1). Los factores personales incluyen las creencias y actitudes que afectan al aprendizaje principalmente con relación a los estímulos conductuales y ambientales. Los factores personales afectan el comportamiento y la forma en que se interpretan las señales del entorno. Esta teoría actualmente forma parte de la Teoría Socio-Cognitiva de Bandura (Bandura, 2001).

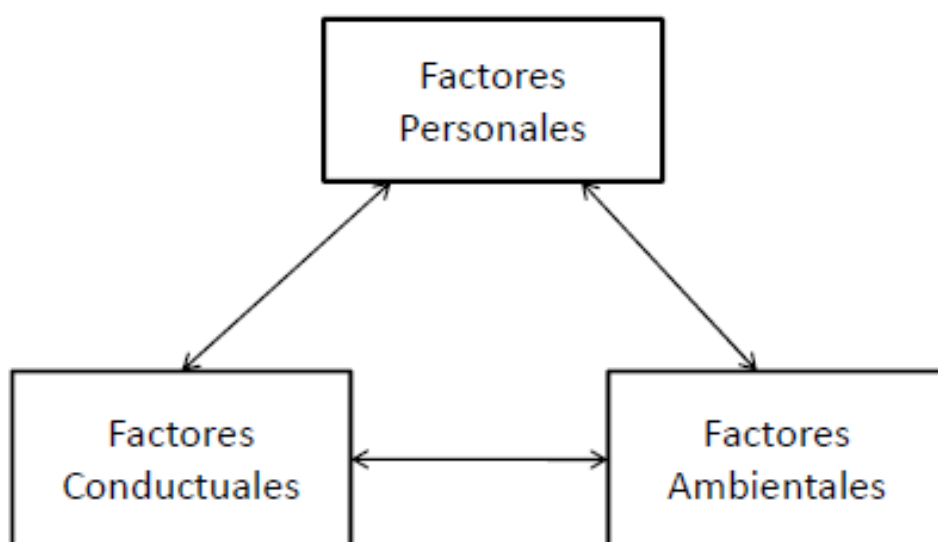


Figura 2.1: Determinismo recíproco adoptado en la teoría del aprendizaje social de Bandura.

En particular, una forma en que los factores personales afectan la interpretación de las señales del entorno es a través de respuestas mediadas. La respuesta a un hecho es mediada por la interpretación cognitiva. Es decir, antes de que se produzca una respuesta, los hechos son interpretados cognitivamente. Frente a un mismo hecho, se pueden producir dos respuestas distintas debido a diferencias en su interpretación cognitiva, que se ve afectada por factores

personales.

Dentro de los factores personales dos variables tienen un papel central en esta teoría: la autoeficacia y expectativa de resultado. La autoeficacia es el grado en que una persona confía en su habilidad para realizar una tarea. La expectativa de resultado es la relación percibida entre realizar una tarea con éxito y obtener un resultado determinado por efecto de esa realización. De acuerdo a esta teoría, variables como la autoeficacia y la expectativa de resultado median la interpretación cognitiva de señales del entorno y los comportamientos (Bandura, 1978a; Schunk y cols., 2014).

Adicionalmente, en esta teoría se identifican varios factores que influyen en el procesamiento de información cognitiva que permite elaborar las percepciones de autoeficacia (Bandura, 1977). Así, un aspecto fundamental en el cambio de la autoeficacia percibida es el procesamiento de información cognitiva derivada de varias fuentes: información enactiva que resulta de realizar una tarea que se está aprendiendo; experiencia vicaria que surge de observar a otro haciendo dicha tarea o conversando con él; información que surge de la persuasión verbal; y, estados fisiológicos.

Otro aspecto importante de esta teoría es el reconocimiento de que la autoeficacia varía según tres dimensiones: magnitud, generalidad e intensidad (Bandura, 1977). Cuanto mayor es la magnitud de la autoeficacia, mayor es la amplitud de dificultades de cierta tarea que comprende. Por otra parte, la generalidad está relacionada con cuánto la autoeficacia para una tarea puede extenderse a otras tareas. Desde otra perspectiva, la generalidad refiere al grado en que cierta experiencia puede producir en un sujeto una percepción de eficacia que va más allá de la tarea específica considerada. Por último, la intensidad está relacionada con el grado en que la autoeficacia puede resistir el efecto negativo de evidencias que sugieren ineficacia. Sujetos que tienen una alta intensidad en sus creencias de autoeficacia se espera que persistan más en sus esfuerzos de afrontamiento a pesar de evidencias que nieguen su eficacia.

Justificación adicional de la hipótesis de investigación.

La teoría del Aprendizaje Social de Bandura, aplicada al proceso de evaluación de la enseñanza, podría implicar que el proceso de respuesta a los SET puede ser mediado cognitivamente por variables como la autoeficacia y las expectativas de resultados. En particular, la autoeficacia y expectativa de

resultado podrían mediar la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza.

Una forma en que esto podría ocurrir es a través del papel de la autoeficacia en la regulación de procesos cognitivos. La autoeficacia tiene un papel regulador en aspectos tales como la calidad del pensamiento analítico, la memoria y las atribuciones causales para éxitos y fracasos, y la orientación a metas (Bandura, 1989, 1991, 1993; Bandura, Barbaranelli, Caprara, y Pastorelli, 1996; Berry, 1999; Schunk, 1990).

Es interesante observar que algunas diferencias individuales en los procesos cognitivos mediados por diferencias en la autoeficacia, podrían también reflejarse en diferencias individuales en el proceso de respuesta a los SET. En parte esto podría estar vinculado a diferencias en autoeficacia para la memoria (Berry, 1999; Berry, West, y Dennehey, 1989) y la autoeficacia para el pensamiento analítico (Bandura y Wood, 1989; Wood y Bandura, 1989). Esto podría explicarse por una relación entre estas variables y la autoeficacia para aprender en un contexto académico. Esta última variable, es la variable directamente considerada en esta investigación.

Por tanto, la autoeficacia para aprender en un contexto académico podría relacionarse con diferencias en la recuperación de la memoria de información de señales utilizadas para evaluar la enseñanza. O bien, podría regular en qué medida se usa una aproximación holística respecto de una aproximación analítica en el proceso de respuesta a los SET. Esto podría ocurrir por un efecto directo de la autoeficacia para aprender en un contexto académico, o bien, indirecto, a través de una posible relación entre la autoeficacia para aprender en un contexto académico y las otras medidas de auto-eficacia mencionadas.

En relación al efecto de la autoeficacia sobre el pensamiento analítico, una menor autoeficacia podría estar relacionada con un “efecto halo” en el proceso de respuesta a los SET, en el que los evaluadores no son capaces de evaluar analíticamente (Bond y Fox, 2015; Engelhard Jr, 2013). Mientras que en relación a las atribuciones causales, la autoeficacia podría afectar los procesos cognitivo-emocionales vinculados a la evaluación de la enseñanza a través de su efecto en las atribuciones realizadas por los evaluadores.

Adicionalmente, estudiantes con una autoeficacia alta en relación a una tarea o actividad de enseñanza tenderían a considerar una tarea más interesante, útil e importante (Wigfield y cols., 1997). Así, la autoeficacia podría estar relacionada positivamente con mayor interés, y una asignación de mayor valor y

utilidad a las tareas de enseñanza propuestas por un docente en un curso.

Por otro parte, la autoeficacia y expectativas de resultados están relacionadas con el proceso de aprendizaje y rendimiento académico (Bandura y cols., 1996; Zimmerman, 1995, 2000). Además, de acuerdo al determinismo recíproco asumido en la Teoría del Aprendizaje Social de Bandura, el proceso de aprendizaje que sigue el estudiante, el proceso de enseñanza y los factores personales de los estudiantes, interactúan y se afectan entre sí. En consecuencia, se espera que la autoeficacia y expectativa de resultado sean capaces de explicar diferencias individuales en cómo la enseñanza facilita el aprendizaje en los estudiantes.

No obstante, de acuerdo a la teoría del Aprendizaje Social de Bandura y a evidencias previas, sería más destacado el papel de la autoeficacia explicando diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes. Esto se debería a una relación más fuerte entre la autoeficacia con el rendimiento académico y el aprendizaje (Shell, Murphy, y Bruning, 1989).

Por ejemplo, estudiantes con una mayor autoeficacia en el contexto de un curso, podrían tener un mejor proceso de aprendizaje, que estudiantes con una menor autoeficacia, en respuesta a cierta propuesta de enseñanza. En parte, esto se podría explicar por comportamientos más adaptativos en estudiantes con mayor autoeficacia, que se caracterizan por comportamientos más marcados de búsqueda de ayuda, participación, persistencia, uso de estrategias, y una mayor orientación a metas de aprendizaje (Bandura y cols., 1996).

En síntesis, de acuerdo a la teoría del aprendizaje social de Bandura y evidencias previas, la autoeficacia y expectativa de resultado estarían vinculadas tanto con la mediación de procesos cognitivo-emocionales relacionados con el proceso de evaluación de la enseñanza, como con diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes. En consecuencia, de acuerdo a los argumentos planteados, y en forma consistente con la hipótesis formulada, se espera que las variables autoeficacia y expectativas de resultados estén implicadas en el proceso de respuesta a los SET.

2.4.3. Teoría de la Atribución de Bernard Weiner.

La teoría de la atribución es una teoría cognitiva que considera las atribuciones causales realizadas por un sujeto, un aspecto clave para explicar la motivación y respuestas emocionales (Weiner, 1985; Weiner y cols., 1979). Una

atribución causal es una explicación causal realizada por un sujeto a un resultado personal. Es decir, las atribuciones causales son interpretaciones sobre las causas percibidas de un resultado personal.

Además, es una teoría cognitiva del comportamiento debido a que considera que existen procesos cognitivos que median entre los estímulos antecedentes y el comportamiento (Weiner, 2013). Desde la perspectiva de esta teoría cognitiva y atribucional tres aspectos son fundamentales: en primer lugar, las causas percibidas de un evento; en segundo lugar, la información que se relaciona con las inferencias causales; y en tercer lugar, las influencias de las atribuciones causales.

Adicionalmente, tiene también relación con otras teorías de la motivación. Al igual que otras teorías que adoptan la posición de *Expectativa × Valor*, considera que la motivación es determinada por lo que uno puede obtener y cuál es la probabilidad de obtenerlo. No obstante, en esta teoría el valor o incentivo de la meta refiere al valor subjetivo de la meta, y no, al valor objetivo que resultaría de las propiedades objetivas de la meta.

Asimismo, el valor subjetivo de la meta tiene una relación isomórfica con su impacto emocional. Si bien, las atribuciones causales no están relacionadas con las propiedades objetivas de las metas, si lo están con el impacto emocional que estas metas producen en el sujeto. Por lo tanto, las atribuciones causales de acuerdo a esta teoría afectan las consecuencias de lograr las metas (Weiner, 1985).

También se pueden señalar algunas similitudes y diferencias con respecto a la Teoría del Aprendizaje Social de Bandura. En la teoría de la atribución los hechos no provocan respuestas afectivas y conductuales en forma directa, sino después de haber sido mediados por algún tipo de interpretación cognitiva. Al igual que en la teoría de Bandura, se considera la mediación cognitiva. No obstante, a diferencia de la Teoría del Aprendizaje Social de Bandura, los juicios atribucionales corresponden a hechos pasados, mientras que los juicios de autoeficacia corresponden a hechos futuros (Graham, 1991).

Asimismo, existe una diferencia entre estas dos teorías en el concepto de expectativas. En la teoría de las atribuciones de Weiner el concepto de expectativa es similar al adoptado en otras teorías de *Expectativa × Valor* (Weiner, 1985). Por el contrario, en la teoría del Aprendizaje Social de Bandura, se distingue entre expectativas de autoeficacia y expectativas de resultados (Bandura, 1978a; Schunk y cols., 2014).

Un aspecto fundamental, distintivo de la teoría de atribuciones causales de Weirner, es el énfasis en procesos cognitivos-emocionales vinculados a las atribuciones causales. En esta teoría, el proceso por el cual un sujeto realiza atribuciones involucra varios componentes: un primer componente de evaluación de resultado (éxito o fracaso); un segundo componente de respuesta atribucional (atribución del resultado a una causa concreta); un tercer componente de respuesta afectiva (respuesta de atribución produce una reacción emocional); y un cuarto componente de reacción conductual (respuesta concreta al resultado).

En particular, el segundo componente, en el cual se atribuye el resultado a una causa concreta, comprende el procesamiento de información de varias fuentes y representa un proceso cognitivo muy complejo (Weiner, 1985). Algunos de los elementos que involucra dicho proceso son: procesamiento cognitivo de la información de la situación actual; recuperación de la memoria de eventos pasados; y, esquemas causales o creencias relativamente estables sobre relaciones entre un evento y las causas percibidas de dicho evento.

Además, cuando un resultado es evaluado como éxito o fracaso, se produce una respuesta emocional general que no depende de la atribución causal: un resultado de éxito en general produce una respuesta de alegría; por el contrario, un resultado de fracaso produce una respuesta de frustración y tristeza. De acuerdo a esta teoría, una vez que se ha dado una respuesta atribucional, y el sujeto atribuye a alguna causa determinada, este experimenta emociones específicas que dependen del tipo de atribución realizada (Schunk y cols., 2014; Weiner, 1985; Weiner y cols., 1979).

Esta teoría considera también que las respuestas atribucionales varían a través de tres dimensiones causales: locus, estabilidad y controlabilidad. La dimensión locus denota si la causa es percibida como interna o externa a la persona. La dimensión estabilidad refiere cuán estable es percibida una causa a través del tiempo. La dimensión controlabilidad denota cuánto el sujeto percibe que tiene control sobre una causa. Las configuraciones específicas de una atribución causal respecto a estas tres dimensiones producen diferentes respuestas afectivas y conductuales (Weiner, 1985, 1986).

De esta forma, en una primera caracterización, podemos considerar el papel que primariamente cumplen cada una de las dimensiones consideradas. Así, la dimensión locus está principalmente relacionado con emociones relacionados con la estima (Graham, 1991; Schunk y cols., 2014; Weiner, 1985). Algunas de las emociones relacionadas con la estima son el orgullo, la confianza y sa-

tisfacción. En particular, el orgullo está relacionado con cómo una persona se siente en relación a sí misma y se produciría cuando una persona atribuye su responsabilidad por producir un resultado de éxito (Weiner, 1986).

Mientras que, la dimensión de estabilidad está principalmente relacionada con el cambio en la expectativa de éxito (Weiner, Nierenberg, y Goldstein, 1976). Adicionalmente, la dimensión estabilidad también está vinculada con la respuesta emocional, específicamente, en relación a la ansiedad experimentada (Arkin y Maruyama, 1979).

Por último, la dimensión controlabilidad está fundamentalmente relacionada con las emociones sociales y juicios interpersonales (Graham, 1991; Schunk y cols., 2014; Weiner, 1986, 2013). Las emociones sociales incluyen la ira, lastima, culpa y gratitud. Mientras tanto, los juicios interpersonales involucran aspectos tales como, decisiones sobre ayudar, evaluación y sentimientos.

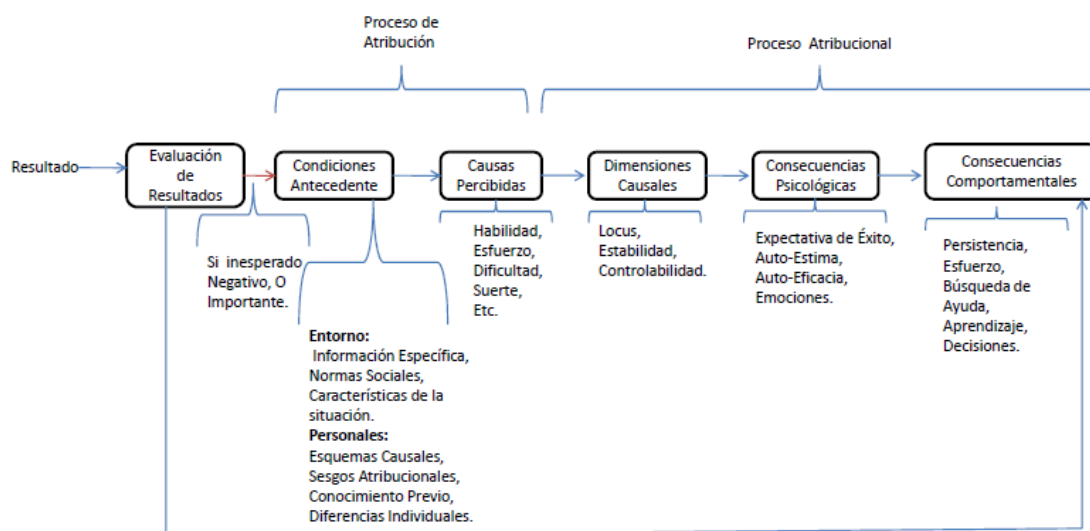


Figura 2.2: Teoría de la Atribución de Bernard Weiner (reelaborado a partir de Weiner (1985))

Así, a partir de las configuraciones específicas de la atribución se esperan diferentes probabilidades de respuestas afectivas. La probabilidad de que se produzca una respuesta afectiva positiva es alta cuando la configuración de la atribución es locus interno, controlable y estable. Una atribución causal que tendría esta configuración es el esfuerzo normal. Para un resultado de éxito, en atribuciones en las que el locus es interno, se espera una probabilidad alta de respuestas afectivas positivas como sentimientos de confianza, satisfacción

y orgullo (Weiner y cols., 1979).

Por otro lado, una respuesta negativa como el enfado, frente a un resultado de fracaso, tiene una probabilidad alta cuando la configuración de la atribución es locus externo (Weiner y cols., 1979). Un ejemplo de atribución causal externa sería el Profesor. Mientras que el atribuir el éxito a la ayuda del profesor es probable que produzca gratitud, atribuir el fracaso a un obstáculo interpuesto por el profesor es probable que produzca enfado (Gredler, 2009).

Una respuesta afectiva negativa como la vergüenza tiene una probabilidad alta, para un resultado de fracaso, cuando la configuración de la atribución es locus interno, incontrolable y estable. Una atribución causal que tendría esta configuración es la aptitud del estudiante. Por otro lado, una respuesta afectiva como la culpa, frente a un resultado de fracaso, tiene una probabilidad alta cuando la configuración de la atribución es locus interno, controlable y estable. Una atribución causal que tendría esta configuración es el esfuerzo normal del sujeto.

En forma similar, a partir de las configuraciones específicas de la atribución se esperan diferentes probabilidades de respuestas conductuales (Graham, 1991; Gredler, 2009; Weiner, 1985). Un aspecto clave en relación con las respuestas conductuales de acuerdo a la teoría de la atribución es que la interpretación de un resultado (atribución causal) determina el tipo de respuesta conductual del individuo.

Un tipo de respuesta conductual importante, que puede ser explicada a través de las configuraciones específicas de las atribuciones, es el comportamiento de afrontamiento a un resultado adverso. Cuando un sujeto atribuye la causa de un fracaso personal a una causa inestable, su expectativa de éxito es mayor que cuando lo atribuye a una causa considerada estable (Weiner y cols., 1976). Por otra parte, su ansiedad es menor cuando atribuye a una causa de un fracaso a una causa inestable, que a una causa considerada estable (Arkin y Maruyama, 1979). Por tanto, la dimensión estabilidad es clave para mantener una mejor expectativa de éxito y menor ansiedad al afrontar un resultado adverso.

Adicionalmente, la persistencia mejora cuando se cambia la atribución causal de baja habilidad a falta de esfuerzo (G. R. Andrews y Debus, 1978), o estrategias pobres Anderson (1983), o bien, a barreras externas temporales (T. D. Wilson y Linville, 1982). Estos cambios estarían, al menos en parte, mediados por cambios en la expectativa de éxito relacionadas con la estabilidad

de las atribuciones (Weiner, 1985).

No obstante, es importante observar que cuando se pasa de una atribución de baja habilidad a una atribución de falta de esfuerzo o de estrategias pobres, no solo hay un cambio en la dimensión estabilidad (que implica menor estabilidad), sino también, un cambio en controlabilidad (que implica mayor controlabilidad). Por tanto, podría haber un efecto también de la dimensión controlabilidad en los cambios observados.

Las atribuciones a causas inestables y controlables como el esfuerzo y uso de estrategias se suele acompañar de un mayor esfuerzo y una revisión o mejora de estrategias cuando se enfrenta un resultado negativo en una tarea cognitiva. Por el contrario, las atribuciones a causas estables e incontrolables como la habilidad en sujetos con teoría de entidad (la habilidad es algo estable que no se puede cambiar) frente a un resultado negativo, suele acompañarse de evitación, falta de persistencia y esfuerzo (Dweck, 1986).

Otro aspecto importante en el afrontamiento de un resultado adverso, que puede explicarse considerando atribuciones causales, es la búsqueda de ayuda (R. Ames y Lau, 1982; Brickman y cols., 1982; Wills y DePaulo, 1991). De acuerdo a Brickman y cols. (1982) una distinción importante es entre la atribución al origen del problema y la atribución a la solución del problema. Un estudiante tendría mayor disposición a buscar ayuda cuando percibe que la solución a su problema esta bajo su control, de tal forma que buscaría el tipo de ayuda que le permita remediar su problema y facilitar su autosuficiencia.

Adicionalmente, R. Ames y Lau (1982) distingue entre patrón de atribuciones relevantes e irrelevantes para la búsqueda de ayuda. Los patrones de atribuciones mediarían en el proceso de tomar la decisión de buscar ayuda o no. Este proceso involucraría el procesamiento de información cognitiva vinculada a varias fuentes: desempeño previo; el patrón atribucional y la aparente disponibilidad y utilidad de los recursos ofrecidos de ayuda. Los estudiantes con un patrón relevante tendrían una mayor probabilidad de buscar ayuda que estudiantes con un patrón irrelevante. Existen varias diferencias, entre ambos patrones atribucionales, que podrían estar vinculadas a una diferente probabilidad de buscar ayuda.

Por un lado, el patrón relevante atribuye un resultado de éxito a la habilidad y el esfuerzo, pero, considera que la causa de su problema es la falta de esfuerzo y no haber comprendido algunos conceptos y principios. Asimismo, estos estudiantes consideran que cuentan con un nivel de habilidad global ne-

cesario. También, perciben que si realizaran un esfuerzo adecuado y recibieran ayuda, las carencias específicas de comprensión se podrían solucionar. Además, no consideran que existan causas externas para su mal desempeño, vinculadas a la dificultad, el docente o la suerte.

Por otro lado, el patrón irrelevante para la búsqueda de ayuda se caracterizaría por atribuciones externas a un mal desempeño. Por tanto, en el patrón irrelevante para la búsqueda de ayuda, sería más probable que el estudiante atribuyera su mal desempeño a causas externas como la dificultad, el docente o la suerte.

Las diferencias que existen entre el patrón relevante e irrelevante para la búsqueda de ayuda podría estar vinculado a un proceso racional de evaluación y toma de decisiones. Si la ayuda se percibe como instrumental y útil para superar un problema, entonces, sería más probable tomar la decisión de buscar ayuda. Si por el contrario, se considera que la ayuda no contribuiría a resolver el problema, entonces se optaría por no buscarla. En este proceso de evaluación y toma de decisiones racional, el tipo de patrón atribucional sería un aspecto clave en el procesamiento de información cognitiva involucrada en la decisión de buscar o no ayuda (R. Ames y Lau, 1982).

Justificación adicional de la hipótesis de investigación.

Lo discutido hasta aquí ayuda a comprender el papel que tienen las atribuciones causales en relación a los dos efectos que componen la hipótesis de investigación propuesta. De acuerdo a esta hipótesis, por un lado, el primer efecto está vinculado a las diferencias individuales en la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza. Por otro lado, la hipótesis plantea un segundo efecto relacionado con formas distintas de aprender y con un efecto diferencial de la enseñanza sobre el aprendizaje de los estudiantes. El segundo efecto planteado, fue analizado con más profundidad y más directamente en la discusión previa. No obstante, tal vez, el segundo efecto requiera una mayor clarificación y discusión.

Por tal motivo, a continuación, se analizarán varios ejemplos concretos desde una perspectiva atribucional, para mostrar con mayor claridad, que la discusión previa es consistente con el primer efecto propuesto en la hipótesis de investigación. Esto es, que la interpretación de las diferencias objetivas que existen en la enseñanza son mediadas cognitivamente por las atribuciones

causales realizadas por el estudiante.

Por ejemplo, dependiendo en parte del patrón atribucional, los estudiantes pueden valorar la ayuda ofrecida como útil para superar los problemas que se les presenten en un curso, o bien, considerarla sin utilidad. Específicamente, los estudiantes con un patrón atribucional relevante para la búsqueda de ayuda, podrían valorar la enseñanza recibida como más útil y valiosa para su aprendizaje. A su vez, estos estudiantes que optarían por buscar ayuda sentirían probablemente gratitud hacia el docente. En consecuencia, la valoración subjetiva de la ayuda suministrada por el docente en el curso sería más positiva para los estudiantes con un patrón atribucional relevante para la ayuda, que los estudiantes con un patrón no relevante.

En forma similar, el tipo de patrón atribucional podría mediar el valor subjetivo atribuido a tareas y actividades propuestas en un curso. Por ejemplo, estudiantes que atribuyen las causas de su desempeño en la tarea a causas con locus interno, inestables y controlables, como el esfuerzo y uso de estrategias, podrían experimentar emociones de orgullo, confianza, competencia y satisfacción, cuando realizan la tarea con éxito. Por el contrario, los estudiantes que atribuyen las causas del desempeño en la tarea a causas externas, estables e incontrolables, como la facilidad de la tarea, no experimentarían estos afectos positivos. Por otra parte, en el caso de un fracaso en la tarea, los estudiantes con una atribución locus interna, inestable y controlable, continuarían realizando esfuerzo, persistirían más en la tarea y revisarían sus estrategias.

Por el contrario, los estudiantes con una configuración externa, estable e incontrolable, al enfrentar un fracaso, responderían con una baja persistencia y esfuerzo, y adoptarían un comportamiento de evitación. Además, experimentarían probablemente una mayor ansiedad, vergüenza, falta de confianza, sentimiento de incompetencia, que los sujetos que tienen una configuración interna, inestable y controlable. Por tanto, la configuración de la atribución locus interno, inestable y controlable se asociaría a una valoración subjetiva más positiva de las tareas y actividades propuestas en un curso, que los estudiantes con la configuración locus externo, estable e incontrolable.

Consideremos también otro ejemplo, cómo afectaría las atribuciones causales a la valoración subjetiva de las señales vinculadas a la claridad del docente. Comparemos cómo sería la diferencia en la valoración subjetiva de la claridad del docente entre, por un lado, una atribución de la comprensión al conocimiento previo del estudiante, su esfuerzo, atención e intención por comprender

(patrón atribucional I), y por otro, una atribución de la comprensión al estilo de comunicación y de organización del contenido presentado por el docente, así como, a la dificultad del contenido seleccionado por el docente (patrón atribucional II).

En el caso de que el estudiante con un patrón atribucional I evalúe su comprensión como mala, este atribuiría su falta de comprensión a sí mismo, pero, con una respuesta adaptativa respecto al problema de comprensión que ha experimentado. En este caso, un estudiante con patrón atribucional I, respondería reforzando su conocimiento previo, aumentando su esfuerzo, atención e intención por comprender. A su vez, una consecuencia lógica de su patrón atribucional, sería que no entendería al docente como su principal causa de falta de comprensión.

Por el contrario, un estudiante con el patrón atribucional II, frente a un problema similar de comprensión, atribuiría su falta de comprensión al estilo de comunicación y organización del docente, y a una excesiva dificultad del contenido seleccionado por este. Además, debido a que el patrón atribucional II corresponde a una configuración externa, estable y no controlable, se esperaría probablemente enojo o ira hacia el docente, que interpondría un obstáculo externo a la comprensión, que además está bajo control del docente.

Adicionalmente, debido a que el patrón atribucional I corresponde a una configuración interna, inestable y controlable, los estudiantes con esta configuración tendrían una respuesta más adaptativa y con una respuesta afectiva más positiva que estudiantes con el patrón atribucional II. Por ejemplo, estudiantes con el patrón atribucional II, que experimenten problemas de comprensión, es probable que presenten ansiedad, indefensión, falta de persistencia y esfuerzo, y comportamientos de evitación. Por el contrario, no se esperaría este tipo de respuesta en los estudiantes con patrón atribucional I. En consecuencia, a partir de las consideraciones realizadas sobre valoración subjetiva de la claridad del docente, se esperaría una valoración más positiva de estudiantes con patrón atribucional I, que de estudiantes con patrón atribucional II.

Los tres ejemplos discutidos, aplicando un análisis atribucional, ilustran y sugieren que los patrones atribucionales explicarían diferencias en la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza. Además, junto a toda la discusión teórica previa, ayudan a clarificar que los argumentos considerados en base a la teoría de la atribución son consistentes con los dos efectos sobre el patrón de respuesta a los SET planteados en la hipótesis de

investigación propuesta. Por tanto, patrones de atribución a causas específicas diferentes, explicarían diferencias individuales en el proceso de respuesta a los SET. En forma similar, atribuciones a causas con diferentes configuraciones en las dimensiones causales, también explicarían diferencias en el proceso de respuesta a los SET.

2.4.4. Teoría de la autodeterminación.

La teoría de la autodeterminación (SDT, por “Self-Determination Theory”) es una aproximación a la motivación y personalidad de los seres humanos que aplica una metateoría organísmica (Deci y Ryan, 2000). Esta aproximación se basa en la asunción de un organismo activo que actúa sobre su ambiente interno y externo. Por tanto, los seres humanos son concebidos como organismos que actúan buscando ser efectivos y para satisfacer sus necesidades. De acuerdo a esta aproximación, en ese proceso de búsqueda, el comportamiento es influenciado por estructuras internas que son elaboradas y refinadas continuamente.

En SDT las necesidades básicas son consideradas universales y constituyen nutrientes psicológicos innatos esenciales para la continuidad del desarrollo, integralidad y bienestar (Deci y Ryan, 2000, 2008). De acuerdo a la SDT, cuando las necesidades básicas son satisfechas, los seres humanos están en condiciones de seguir sus tendencias básicas hacia el desarrollo, integración y bienestar. Por el contrario, cuando estas necesidades no son satisfechas, los seres humanos tienen una tendencia hacia problemas de salud mental y malestar.

Estas necesidades fueron introducidas inductivamente a partir de investigación empírica, siguiendo el principio de parsimonia. Por tanto, en la definición de las necesidades básicas se buscó definir el menor número posible de necesidades básicas que fueran capaces de explicar el mayor número posible de fenómenos (Vansteenkiste, Niemiec, y Soenens, 2010). Así, la SDT sostiene que para explicar la motivación humana y la personalidad es necesario considerar tres necesidades básicas: autonomía, competencia y relacionamiento.

La competencia es la necesidad del ser humano de percibir que es capaz de actuar en forma competente y eficaz. Este concepto se superpone parcialmente con el concepto de autoeficacia considerada en la teoría de Aprendizaje Social de Bandura (Bandura, 1977, 1982; Bandura y Walters, 1977). No obstante, mientras en la SDT se considera la competencia como una necesidad básica, en

la teoría de Bandura la autoeficacia no es conceptualizada como una necesidad psicológica. Sin embargo, en ambas teorías estos constructos están relacionados con la auto percepción de eficacia o competencia y tienen una relación positiva con una serie de resultados personales.

Por otra parte, la autonomía refiere a la necesidad del ser humano de percibir que tiene elecciones y control sobre sus propias acciones. Desde una perspectiva atribucional se relaciona con la percepción de un locus interno de causalidad (R. M. Ryan y Deci, 2000b). En la teoría de la atribución de Bernard Weiner el locus de causalidad es un constructo fundamental (Weiner, 1985) primariamente relacionado con las emociones vinculadas a la estima. Por tanto, al igual que en el caso de la necesidad de competencia, se aprecia una relación entre la necesidad de autonomía y un constructo importante de otra teoría socio cognitiva.

A su vez, el término relacionamiento refiere a la necesidad del ser humano de percibir que está conectado a otros seres humanos a través de relaciones positivas (Deci y Ryan, 2002; Vansteenkiste y cols., 2010). Esta necesidad está vinculada al sentimiento de querer y ser querido por otros. También designa la necesidad de pertenecer a un grupo, por lo que en ocasiones se la denomina necesidad de pertenencia (Baumeister y Leary, 1995).

La postulación de la existencia de necesidades básicas innatas y universales realizada en la SDT constituye un principio unificador. La SDT es una macro-teoría organizada en varias mini-teorías, incluyendo: la teoría de la evaluación cognitiva y la teoría de la integración orgánica. En cada una de estas mini-teorías la dialéctica orgánica y las necesidades básicas postuladas contribuyen a la coherencia que caracteriza a la SDT (Vansteenkiste y cols., 2010).

Teoría de Evaluación Cognitiva.

La primer mini-teoría que forma parte de la SDT es la teoría de evaluación cognitiva (o CET, por “Cognitive Evaluation Theory”). Esta teoría se originó a partir de investigaciones relacionadas con el efecto de recompensas e incentivos externos en la motivación de sujetos que realizan tareas interesantes y disfrutables (Deci, 1971, 1972; Deci y Cascio, 1972; Deci y cols., 1973; M. Lepper y Greene, s.f.; M. R. Lepper, Greene, y Nisbett, 1973).

En la CET la variable respuesta central es la motivación intrínseca. Una

primera definición de carácter operativo, adoptada en investigaciones empíricas tempranas, definen que un comportamiento es intrínsecamente motivado cuando se realiza sin que exista una recompensa externa aparente (Deci, 1971, 1972; Deci y Cascio, 1972; Deci y cols., 1973). A partir de esta definición, en estos trabajos se elaboró una operacionalización primaria de la medida de motivación intrínseca como el porcentaje del tiempo dedicado a una actividad en condiciones de elección libre entre esa y otras actividades disponibles.

La anterior definición operativa, se complementa con el reconocimiento de que la recompensa recibida por el sujeto, motivado intrínsecamente para llevar a cabo una tarea, está relacionada con su respuesta afectiva a realizar dicha tarea. Por ello, una operacionalización secundaria utilizada por Deci (1971) fue el auto-reporte del nivel de interés y disfrute. De esta forma, cuando la motivación para realizar una actividad es intrínseca, la recompensa surge de realizar la propia actividad. Dicha recompensa está relacionada con la satisfacción, interés y disfrute que representa su realización. Así, en este caso la recompensa está en la propia actividad y no en una recompensa externa contingente a su realización. No obstante, de acuerdo a la CET, la motivación intrínseca no implica la búsqueda activa del disfrute, sino, que el disfrute surge como un subproducto de la completa inmersión en la actividad (R. M. Ryan y Deci, 2000a; Vansteenkiste y cols., 2010).

Por el contrario, cuando la motivación para realizar una tarea es extrínseca, esta motivación surge de una recompensa externa a la tarea que se obtiene como resultado de realizar la misma. Esta distinción entre motivación intrínseca y extrínseca puede mirarse desde una perspectiva atribucional. Así, el locus de causalidad, en relación a la participación del sujeto en la actividad, es interno para un sujeto intrínsecamente motivado. Por el contrario, para un sujeto extrínsecamente motivado puede presentar diferentes grados de percepción del locus de causalidad (R. M. Ryan y Deci, 2000a). Es decir, un sujeto que participa en una tarea para la cual su motivación es extrínseca, puede variar su percepción del locus de causalidad de completamente externo a completamente interno.

Otro aspecto importante que ayuda a entender la naturaleza de la motivación intrínseca, y su relación con la motivación extrínseca, está relacionada con la pregunta de si estos constructos son extremos opuestos de un continuo bipolar, o bien, si son dos constructos unipolares independientes. Estudios sobre el desarrollo de la motivación intrínseca y extrínseca en niños aportaron

evidencia consistente con la independencia y unipolaridad de estos constructos (M. R. Lepper, Sethi, Dialdin, y Drake, 1997; Schunk y cols., 2014).

Por otro lado, de acuerdo a la CET la base psicológica de la motivación intrínseca esta en las necesidades de la persona de sentirse competente y auto-determinada (Deci, 1975b). Para que una persona este intrínsecamente motivada para realizar una actividad, esta debe percibir que la realización de la misma satisface sus necesidades de competencia y autonomía. Solamente satisfacer una de ellas no es suficiente.

Por otra parte, si bien se ha identificado a las necesidades de competencia y autonomía como dos factores muy importantes que afectan la motivación intrínseca, la necesidad de relacionamiento tiene un papel también en la motivación intrínseca, si bien, este papel sería de carácter más distal (Deci y Ryan, 2000). Esto es, la necesidad de relacionamiento actuaría como telón de fondo en relación a la motivación intrínseca percibida por el sujeto. Es decir, para que el sujeto pueda percibir que está intrínsecamente motivado para realizar una actividad, este debe percibir en su vida suficiente satisfacción de la necesidad de relacionamiento. O bien, experimentar satisfacción de la necesidad de relacionamiento en el contexto de realización de la actividad.

Por tanto, aplicando este principio básico de la CET, que relaciona la motivación intrínseca con la percepción de satisfacción de las necesidades básicas, es posible explicar el efecto de eventos externos (recompensas y retroalimentaciones) sobre la motivación intrínseca (R. M. Ryan, 1982). Para ello, es necesario considerar como características de dicho evento afectan la percepción de satisfacción de las necesidades básicas.

De acuerdo a la CET, los eventos externos tienen dos aspectos que explican su efecto sobre la motivación intrínseca (Deci y Ryan, 1980; R. M. Ryan, 1982): un aspecto informativo y un aspecto controlador. El aspecto informativo de un evento externo aporta información relativa a la competencia en la actividad realizada. Por otra parte, el aspecto controlador del evento externo aporta señales sobre el carácter controlador del mismo, esto es, que contribuyen a que el evento externo sea percibido como una presión para obtener un resultado comportamental determinado.

Así, de acuerdo a la CET, un evento externo tendría un efecto positivo en la motivación intrínseca cuando se verifican las siguientes condiciones: primero, la parte informativa resulta más saliente en la percepción de la persona, que la parte controladora; segundo, la parte informativa del evento aporta infor-

mación que incrementa la percepción de competencia. Dada estas condiciones, se promueve la percepción de satisfacción de la necesidad de competencia y de autonomía. En otros términos, la percepción de la persona es afectada por el evento externo, de tal forma que, este incrementa su percepción de competencia, al mismo tiempo que, no percibe una presión o control que afectaría negativamente su autonomía.

Modelo cognitivo del comportamiento orientado a metas.

En este punto, se considera importante precisar la aproximación cognitiva a la motivación, considerada por la SDT, que es destallada por Deci (1975a). La percepción de una posible satisfacción de las necesidades de competencia y autonomía constituyen en este caso los motivos. Estos motivos consisten en representaciones mentales de estados futuros. La percepción de una satisfacción potencial de las necesidades aporta la energía necesaria para seleccionar metas y comportarse de formas que la persona considera que podrían conducirla a satisfacer sus necesidades. Las elecciones de los comportamientos a ser llevados a cabo se basan en las expectativas de los estados finales a los que conducirían dichos comportamientos.

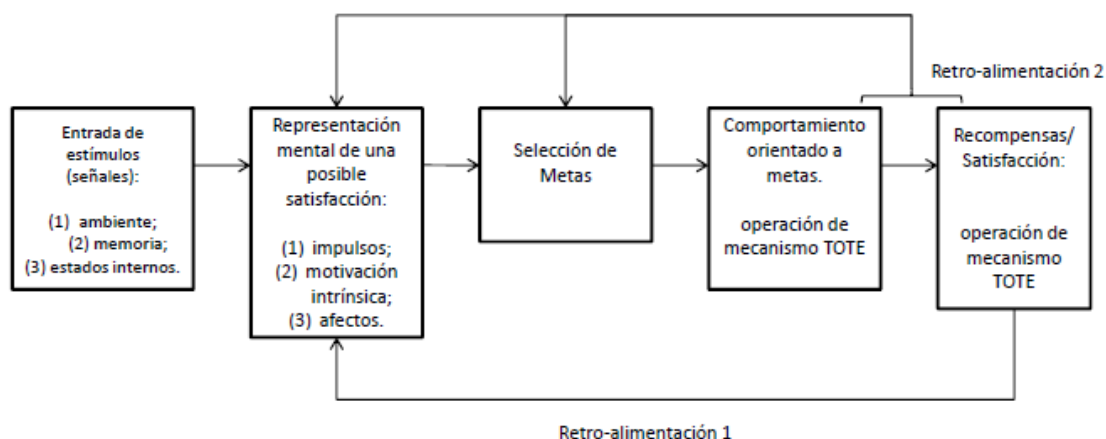


Figura 2.3: Modelo cognitivo del comportamiento motivado (reelaborado a partir de Deci (1975a))

En la Figura 2.3 se presenta un esquema general de la aproximación cognitiva al comportamiento motivado adoptado por la SDT (Deci, 1975a). La primera etapa de procesamiento de información comienza con los estímulos de

entradas o señales que aportan información al sistema. Estos estímulos pueden provenir del ambiente, de la memoria, o bien, de estados internos del sistema.

En la segunda etapa, estas señales son utilizadas para elaborar las representaciones mentales de posibles satisfacciones. Un tipo de representación mental corresponde a posibles satisfacciones de las necesidades intrínsecas, de competencia y autonomía. No obstante, también se consideran en el modelo las representaciones mentales de impulsos, por ejemplo vinculadas a necesidades de alimentación. Adicionalmente, el modelo considera también los afectos. En esta etapa de procesamiento, podría hacerse prominente una representación mental de una posible satisfacción que correspondería a un estado emocional positivo o menos negativo.

En la tercera etapa, las representaciones mentales de las posibles satisfacciones conducen a la selección de metas. Así, a partir de las representaciones mentales de las posibles satisfacciones se seleccionan metas que se esperan conduzcan a la satisfacción, correspondiente a la representación mental elaborada en la etapa previa.

En la cuarta etapa, las metas seleccionadas en la etapa anterior, orientan el comportamiento. Esto es, le dan dirección y cuando las metas son alcanzadas el comportamiento finaliza. En el modelo se propone que en esta etapa funciona un mecanismo TOTE (por, “Test, Operate, Test, Exit”) descrito por [Miller, Galanter, y Pribram \(1960\)](#). Si el comportamiento no conduce a la obtención de la meta, existen en el modelo otros mecanismos de terminación de esta etapa. Estos mecanismos conducen a una modificación del valor de la meta, que conduce a la selección de otra meta. A su vez, esto conduce a otros comportamientos orientados a la nueva meta adoptada.

En la quinta etapa, si los comportamientos orientados a metas, condujeron a recompensas (extrínsecas, intrínsecas o afectivas) que concuerdan con las representaciones mentales de las satisfacciones esperadas, entonces, la prominencia de las representaciones de las necesidades disminuye. Si no fuera así, la representación mental de posibles satisfacciones conduciría a una nueva secuencia de comportamiento dirigido a metas.

Este modelo cognitivo del comportamiento orientado a metas, nos ayuda a entender más el comportamiento intrínsecamente motivado y cómo este se relaciona con un ambiente cambiante e inestable. Por ejemplo, si una persona tiene un estado emocional relativamente positivo, y no se hace prominente ningún impulso (por ejemplo, no presenta hambre o sed), la persona sigue un

comportamiento orientado a la satisfacción de sus necesidades de competencia y autonomía.

De esta forma, dicha persona, busca desafíos, sigue comportamientos de exploración, busca ampliar sus experiencias, poner a prueba sus límites y recibe con apertura lo novedoso. Con esto sigue un camino de crecimiento, aprendizaje y creatividad. A su vez, estos comportamientos producen satisfacción de necesidades básicas de competencia y autonomía, que conducen a una respuesta afectiva de satisfacción, interés y disfrute, que energiza y dirige este comportamiento intrínsecamente motivado. Este proceso, eventualmente conduce a generar nuevas metas orientadas a la satisfacción de necesidades intrínsecas, que llevan a experimentar nuevos desafíos y aprendizajes.

Este proceso continuo puede ser interrumpido transitoriamente cuando cambios en el ambiente externo o en estados internos hacen prominente un impulso o necesidad afectiva. Por ejemplo, un niño que permaneció mucho tiempo inmerso en un juego que le produce gran interés y disfrute, puede experimentar hambre y sentir un olor de una comida deliciosa, lo que puede conducir a una representación mental de una posible satisfacción del hambre que siente este niño. Esto puede continuarse con la selección de una meta de alimentación y un comportamiento orientado por dicha meta. Una vez que el niño satisface este impulso (vinculado con el hambre), las representaciones mentales relacionadas con la satisfacción de necesidades intrínsecas de competencia y autonomía se hacen nuevamente prominentes, y esto, conlleva a retomar comportamientos intrínsecamente motivados.

Teoría de Integración Organísmica.

El modelo cognitivo del comportamiento orientado a metas, descrito en la sección anterior, pone en perspectiva la motivación intrínseca, que es el foco de la CET. También, explícitamente considera otros aspectos relevantes del comportamiento motivado. En particular, describe cómo el comportamiento orientado a metas de una persona se ajusta en forma continua a un ambiente incierto y cambiante.

Así, este modelo explica cómo se articulan en forma continua la satisfacción de la motivación intrínseca, con otros aspectos del comportamiento que son extrínsecamente motivados. Estos comportamientos extrínsecamente motivados, en tanto, se realizan por razones instrumentales, para obtener algún

resultado separable del comportamiento en sí, que podría ayudar a ajustarse a las demandas del ambiente.

No obstante, este modelo no explica cómo la motivación extrínseca puede jugar un papel en las tendencias orgánsmicas al crecimiento e integración. La teoría de la integración orgánsmica (OIT, por “Organismic Integration Theory”) se focaliza en la motivación extrínseca y en el proceso de internalización, que es una manifestación de las tendencias del organismo al crecimiento e integración, vinculada con la internalización de la regulación de comportamientos extrínsecamente motivados (Deci y Ryan, 1985; R. M. Ryan y Deci, 2000a, 2000b, 2017).

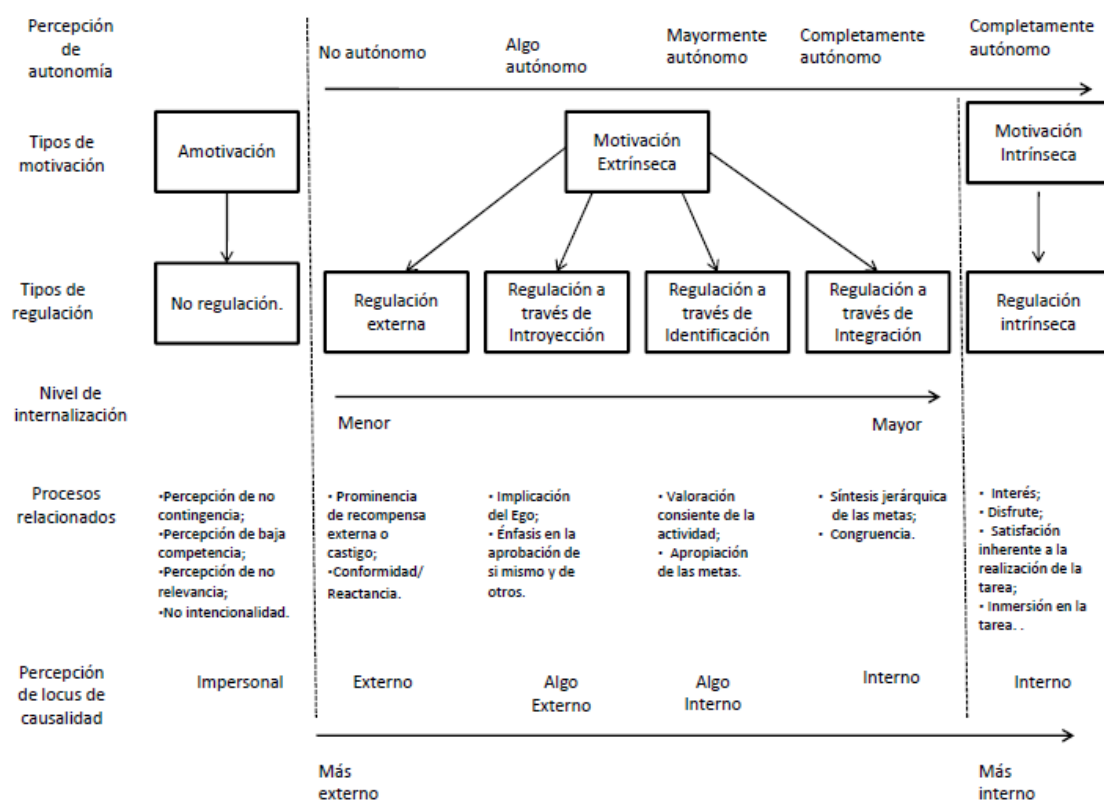


Figura 2.4: Esquema general de la teoría de integración orgánsmica (reelaborado a partir de R. M. Ryan y Deci (2017))

En la Figura 2.4 se presenta un esquema general de la OIT. Esta teoría distingue entre tres tipos de motivación: amotivación, motivación extrínseca y motivación intrínseca. A su vez, se proponen diferentes mecanismos de regulación vinculados con los tres tipos básicos de motivación. En particular, los mecanismos de regulación de la motivación extrínseca, reflejan diferencias

en la forma de internalización de la regulación extrínseca (R. M. Ryan y Deci, 2000a, 2017). Además, cada uno de estos mecanismos de regulación de la motivación se relaciona con diferencias fenomenológicas (R. M. Ryan y Deci, 2017). Adicionalmente, cada tipo de regulación se diferencia en la percepción del locus de causalidad y en la percepción de autonomía (R. M. Ryan y Connell, 1989; R. M. Ryan y Deci, 2017).

Así, la primera distinción importante establecida en la OIT son los tipos de motivación considerados. La Amotivación es un estado en el cual la persona no está motivada a llevar a cabo un comportamiento, o bien, el comportamiento no está mediado por intencionalidad. La falta de intencionalidad representa una diferencia fundamental con los comportamientos motivados. Esta diferencia substancial se representa en la Figura 2.4 con una línea vertical que separa la amotivación de los otros tipos de motivación. Una persona amotivada percibe que no existe contingencia entre la realización del comportamiento y un resultado esperado, o bien, percibe una baja competencia para realizar efectivamente el comportamiento. Una persona en un estado de amotivación no responde a un control externo, ni percibe algún grado de autonomía.

Por otra parte, la motivación extrínseca es un tipo de motivación que implica llevar a cabo una actividad por un resultado separable de la actividad en sí (R. M. Ryan y Deci, 2000b). Esto contrasta con la motivación intrínseca o necesidad de competencia y autonomía, con la cual se lleva a cabo una actividad sin que este implicada una recompensa externa aparente. Más bien, la persona intrínsecamente motivada para realizar una actividad experimenta satisfacción de sus necesidades de competencia y autonomía (Deci, 1975a; R. M. Ryan y Deci, 2000b; Vansteenkiste y cols., 2010). En consecuencia, siente interés, disfrute, e inmersión en la actividad que realiza.

En segundo lugar, otro tipo de distinción realizada por la OIT es entre diferentes tipos o estilos de regulación de la motivación (R. M. Ryan y Deci, 2000a, 2017). En relación a la motivación extrínseca, se distinguen cuatro tipos de regulación: regulación externa; regulación a través de introyección; regulación a través de identificación; y, regulación a través de integración. Estos estilos de regulación se diferencian en el nivel de internalización que producen, y el nivel de percepción del locus de causalidad y de autonomía. En el contexto de la OIT el término internalización refiere al proceso de adoptar valores, creencias, o regulaciones comportamentales, a partir de fuentes externas y transformarlas en propias (R. M. Ryan, Connell, y Deci, 1985; R. M. Ryan y Deci, 2000a,

2017).

Por tanto, podemos distinguir entre los tipos de regulación de la motivación extrínseca, por el nivel y profundidad de la internalización que implican. En la regulación externa el comportamiento es motivado por y dependiente de contingencias de recompensas o castigos. En este caso la percepción del locus de causalidad es completamente externo y el comportamiento realizado se percibe como no autónomo o controlado. Así, en este tipo de regulación el nivel de internalización alcanzado corresponde con un comportamiento de aceptación pasiva u obediencia, donde se acepta realizar un comportamiento, a cambio de una recompensa, o bien, evitar un castigo. Si bien, este es un nivel muy bajo de internalización, es mayor al que se produce en personas amotivadas donde la persona es renuente y no tiene intención de llevar a cabo el comportamiento (R. M. Ryan y Deci, 2000a).

Un nivel de internalización superior a los anteriores, se alcanza en la regulación introyectada. Este tipo de regulación se basa en contingencias afectivas y evaluativas dentro del individuo de las que depende el auto-valor y auto-estima. Desde un punto de vista fenomenológico, se relaciona con experimentar un sentimiento de deber u obligación de hacer algo, o bien, enfrentar un sentimiento de ansiedad y auto-desprecio (R. M. Ryan y Deci, 2017). La regulación introyectada se basa en parte en un mecanismo de proyección. En el contexto de la OIT, el término proyección refiere a la atribución de auto-aprobaciones y auto-reprobaciones a otros, imaginando que estos otros aprobaran o desaprobaran en función de ciertos comportamientos o resultados.

El siguiente nivel de profundidad de la internalización, se produce en la regulación a través de identificación. La identificación es un tipo de internalización definida por una aprobación consciente de valores y regulaciones (R. M. Ryan y Deci, 2000a). Un sujeto se identifica con un valor o regulación cuando considera que son de valor e importancia personal. Por tanto, cuando la regulación es identificada, la persona considera personalmente valioso o importante llevar a cabo cierto comportamiento. Este comportamiento es visto por la persona como instrumental para lograr algo personalmente importante. Pero, no lo hace por obligación, como en la regulación introyectada, o bien, por un control externo, como ocurre en la regulación externa.

Por último, el mayor nivel de internalización se produce en la regulación integrada, que es un tipo de regulación basada en la integración. La integración es un tipo de internalización definida por poner valores y regulaciones

internalizados en congruencia con otros aspectos de uno mismo (R. M. Ryan y Deci, 2000a). La integración es un proceso que requiere auto-reflexión y asimilación recíproca. El proceso de integración implica poner en congruencia una internalización con las necesidades básicas y con otras internalizaciones. Esta mayor congruencia permite una mayor percepción de auto-determinación, que la percibida en la regulación identificada.

Teoría de Necesidades Psicológicas Básicas.

Gran parte de los aspectos más relevantes de la SDT, para la presente investigación, fueron discutidos en las secciones previas. No obstante, quedan algunos puntos relevantes aún no discutidos. Estos, están vinculados a dos de las restantes mini-teorías de la SDT.

El primer punto está relacionados con la mini-teoría de Necesidades Psicológicas Básicas (BNPT, por “Basic Psychological Needs Theory”). Esta teoría especifica los nutrientes psicológicos innatos necesarios para satisfacer las necesidades psicológicas básicas (R. M. Ryan y Deci, 2017; Vansteenkiste y cols., 2010). Así, además de considerar estas necesidades, considera también las dimensiones del ambiente social que apoyan o frustran la satisfacción de las necesidades básicas. La BPNT actualmente distingue tres dimensiones del ambiente social: soporte a la autonomía, que refiere a las características que sirven de apoyo a la autonomía; estructura, es decir una estructuración adecuada relacionada con el apoyo a la competencia; y compromiso, relacionado con la calidez y receptividad del ambiente social, vinculada con el apoyo al relacionamiento.

Así, de acuerdo a Vansteenkiste y cols. (2010) un docente apoya la autonomía del estudiante cuando: trata de entender su perspectiva; ofrece a este la posibilidad de tomar decisiones o elecciones significativas para él y su aprendizaje; y por último, ofrece una base lógica significativa y realista cuando las opciones son restringidas. Lo contrario a apoyar la autonomía, es tener una aproximación controladora, que significa intentar controlar el pensamiento, las emociones y comportamiento de los estudiantes. Esto es, una orientación controladora frustra su necesidad de autonomía.

Por otra parte, en la BPNT un aspecto que se destaca respecto al apoyo a la competencia es proveer de estructura (Reeve, 2006). Un ambiente de aprendizaje es bien estructurado cuando es consistente y predecible. Por ejemplo,

proveer de una estructura adecuada es dejar claro qué se espera de los estudiantes y cuáles son las consecuencias de satisfacer o no dichas expectativas. Lo contrario a esto, la falta de estructura, tiene el efecto de frustrar la necesidad de competencia.

Por su parte, el apoyo al relacionamiento se ha vinculado con una dimensión del ambiente social denominada compromiso (Skinner y Belmont, 1993). Un docente es comprometido en la medida que: dedica tiempo y recursos a los estudiantes; es afectuoso y muestra disfrutar la relación con estos; y, es sensible a la perspectiva de estos (sus intereses, diferentes grados de participación, respuestas emocionales). El compromiso con los estudiantes se considera opuesto a la expresión de descuido y rechazo por estos, que tiene el efecto de frustrar la necesidad de relacionamiento.

A partir de la discusión previa sobre la BPNT, emergen algunas observaciones relevantes para aplicar una aproximación basada en la SDT a la evaluación de la hipótesis central en esta investigación.

En primer lugar, desde la perspectiva de la BPNT, ciertas características de la enseñanza pueden apoyar, o bien, frustrar las necesidades psicológicas básicas. En particular, esta teoría identifica y destaca tres dimensiones del ambiente social: soporte a la autonomía, estructura y compromiso. Por tanto, una primera observación es que una forma de valorar la calidad de la enseñanza, es considerar en que medida y proporción, ésta ofrece apoyo a la autonomía, una estructura adecuada y compromiso.

En segundo lugar, la BPNT especifica como la satisfacción o frustración de las necesidades básicas se relaciona con el bienestar y un funcionamiento completo. Así, una segunda observación es que esta teoría aporta un aproximación a cómo la satisfacción de las necesidades psicológicas básicas afectan el aprendizaje del estudiante y su bienestar emocional. Mientras que la OIT se basa en una aproximación que se enfoca en la relación entre la satisfacción de las necesidades básicas con las formas de regulación de la motivación y con el proceso de internalización; la BPNT introduce una perspectiva sistemática y profunda que considera las dimensiones del ambiente social y su relación con la satisfacción o frustración de las necesidades básicas y su impacto en el bienestar y funcionamiento del organismo.

En tercer lugar, la BPNT nos ayuda a comprender mejor las evaluaciones o valoraciones que realizan los estudiantes y su función en la interacción entre estos y el ambiente social. En particular, una tercera observación es que la

BPNT nos puede ayudar a comprender mejor el papel y relación entre la percepción de satisfacción de las necesidades básicas y de las valoraciones de la enseñanza, y cómo estas percepciones se relacionan con el aprendizaje y bienestar del estudiante.

Teoría de Contenidos de Metas.

El último punto pendiente, relacionado con la evaluación de la hipótesis central de esta investigación, a partir de una aproximación sustentada en la SDT, está relacionados con la mini-teoría de contenidos de metas (GCT, por “Goal Contents Theory”). Esta teoría distingue entre dos tipos de metas: metas intrínsecas (MI) y metas extrínsecas (ME). Estos tipos de metas, de acuerdo a la GCT, están relacionadas en forma diferente con la satisfacción de las necesidades básicas (Kasser y Ryan, 1996; R. M. Ryan y Deci, 2017). Mientras que la búsqueda y logro de metas intrínsecas está asociado a la satisfacción de necesidades psicológicas básicas, la búsqueda y logro de metas extrínsecas no lo está, e incluso puede estar asociado con la frustración de las necesidades psicológicas básicas.

Las metas intrínsecas son metas inherentemente valiosas o satisfactorias para el individuo (Kasser y Ryan, 1996). Algunos ejemplos de metas intrínsecas son: la contribución al medio social, la salud, el desarrollo personal, el aprendizaje y filiación. Por otra parte, de acuerdo a Kasser y Ryan (1996) el valor de las metas extrínsecas son dependientes de evaluaciones de otros. La fama, el éxito, el dinero y la apariencia personal, son algunos ejemplos de metas extrínsecas. Por ejemplo, si consideramos metas vinculadas específicamente con un curso universitario, mientras que lograr aprender los contenidos de un curso corresponde a una meta intrínseca, el obtener una buena calificación en dicho curso corresponde a una meta extrínseca.

Es importante observar que los tipos de metas intrínsecas y extrínsecas, consideradas en la GCT, difieren de los tipos de regulación intrínseca y extrínseca, referidos en la OIT. Mientras que tipos de metas se refieren al “para qué”, que está relacionado con cuál es el objetivo de la actividad; los tipos de regulación de la motivación están relacionadas con “el por qué”, que está vinculado con las razones que lo llevan al esfuerzo para conseguir dicho objetivo (Deci y Ryan, 2000).

La distinción entre metas intrínsecas y extrínsecas realizada por la GCT

tiene relevancia para explicar diferencias en el aprendizaje, desempeño y persistencia de los estudiantes. De acuerdo a [Vansteenkiste, Lens, y Deci \(2006\)](#), una conclusión emergente de varios experimentos de campo que consideran el efecto del contenido de metas, es que la adopción de metas intrínsecas está relacionada con un compromiso más profundo en las actividades de enseñanza, mejor aprendizaje conceptual, y mayor persistencia en actividades de aprendizaje, respecto de la adopción de metas extrínsecas. Adicionalmente, [Vansteenkiste, Simons, Lens, Sheldon, y Deci \(2004\)](#) aportan evidencias que muestran que este efecto de las metas intrínsecas, además, tiene una interacción positiva con la exposición a un contexto de apoyo a la autonomía. Así, de acuerdo a [Vansteenkiste y cols. \(2004\)](#) la adopción de metas intrínsecas y el apoyo a la autonomía tendrían ambos por separado un efecto positivo, pero, en conjunto tendrían un efecto sinérgico.

Adicionalmente, [Svanum y Aigner \(2011\)](#) aportaron evidencias que sugieren que el contenido de las metas adoptadas por los estudiantes está relacionada con los puntajes estudiantiles de la enseñanza. De acuerdo a [Svanum y Aigner \(2011\)](#) los estudiantes que adoptan metas intrínsecas evaluarían más positivamente la enseñanza que estudiantes que adoptan metas extrínsecas. No obstante, dicha investigación tiene algunas limitaciones importantes, analíticas e instrumentales. Por ejemplo, cada una de las dimensiones la efectividad en la enseñanza consideradas por [Svanum y Aigner \(2011\)](#) es medida mediante un único ítem. Adicionalmente, la aproximación analítica utilizada por [Svanum y Aigner \(2011\)](#) no permite aislar la contribuciones de cada uno de los elementos básicos que afectan el proceso de respuesta, esto es, efectividad del docente, severidad del estudiante y dificultad del ítem considerado.

2.4.5. Modelo de Dweck y Leggett.

Dweck y Leggett ([Dweck, 1986, 2000; Dweck y Leggett, 1988](#)) propusieron un modelo sobre los procesos motivacionales y como estos afectan el logro académico y aprendizaje, que aporta otra perspectiva útil para evaluar nuestra hipótesis de investigación. En este modelo una de las variables centrales es el tipo de metas. Un primer tipo de metas se denominan metas de aprendizaje, donde el sujeto busca desarrollar competencia o perfeccionar la ejecución de una tarea. Cuando un sujeto adopta metas de aprendizaje su respuesta comportamental es aprender, o dominar una tarea, ya sea su capacidad auto-percibida

alta o baja.

Un segundo tipo de metas son las metas de ejecución, donde el sujeto procura demostrar a sí o a otras personas su competencia, o bien, evitar poner en evidencia la falta de competencia. Cuando un sujeto adopta metas de ejecución, su respuesta comportamental varía dependiendo de su capacidad auto-percibida. Si la percepción sobre su propia capacidad es alta, su respuesta comportamental es demostrar competencia. Por otra parte, si su percepción sobre su propia capacidad es baja, su respuesta comportamental es evitar demostrar falta de competencia.

De acuerdo a este modelo un aspecto clave son las auto-teorías del sujeto. En particular, las creencias acerca de la naturaleza fija o maleable de su inteligencia o aptitud (Teoría de la Entidad/Teoría Incremental). Los sujetos con una visión de entidad sobre la inteligencia creen que esta (aptitud o habilidad) es fija y no hay nada que puedan hacer para mejorarla. Por el contrario, sujetos que tienen una perspectiva incremental creen que con esfuerzo pueden mejorar su inteligencia o aptitud.

En el modelo de Dweck y Leggett los sujetos con una perspectiva incremental tienden a adoptar metas de aprendizaje. Así, estos sujetos quieren desarrollar su competencia y se interesan por aprender. Por el contrario, los que poseen una perspectiva de entidad tienden a adoptar metas de ejecución. En este caso, estos sujetos quieren demostrar su competencia, o bien, intentan evitar evidenciar falta de competencia. Mientras que los sujetos con una visión incremental pueden preferir tareas que representen ciertos desafíos, los que tienen una visión de entidad prefieren tareas que les permitan lucir capaces o inteligentes.

Adicionalmente, las auto-teorías de los sujetos son también importantes en este modelo para explicar creencias sobre el esfuerzo y el tipo de atribuciones causales realizadas a los fracasos. Los sujetos que tienen una perspectiva de entidad relacionan el ejercer esfuerzo con la falta de inteligencia o aptitud. Por el contrario, los sujetos que tienen una perspectiva incremental ven el esfuerzo como un medio para aprender y desarrollar la inteligencia y la aptitud.

En forma similar, las atribuciones causales a los fracasos son también diferentes entre sujetos con una visión de entidad respecto de otros con una perspectiva incremental. Para un sujeto con una visión de entidad los errores o fracasos son interpretados como una consecuencia de la falta de habilidad. Por el contrario, para los que tienen una perspectiva incremental, los errores

o fracasos son atribuidos a la falta de esfuerzo o estrategias adecuadas.

Estos dos tipos de auto-teorías a su vez, se relacionan con patrones motivacionales distintos (Dweck, 1986). El patrón adaptativo se caracteriza por la búsqueda de desafíos y superación, así como, la persistencia frente a la presencia de obstáculos. Los sujetos que presentan este patrón adaptativo, muestran además disfrutar el esfuerzo implicado en la búsqueda del dominio de una tarea. Por el contrario, el patrón maladaptativo se caracteriza por la evitación de desafíos y baja persistencia frente a dificultades. Mientras que los sujetos con una visión incremental tienden a adoptar patrones motivacionales adaptativos independientemente de la percepción de su propia competencia, los sujetos con teorías de entidad, adoptan un patrón motivacional adaptativo solo cuando evalúan su propia competencia como elevada.

El modelo de Dweck y Leggett plantea que el tipo de auto-teoría (Entidad vs. Incremental) se relaciona con el tipo de meta (Aprendizaje vs Ejecución), mientras que el tipo de meta se relaciona con el tipo de patrón motivacional (Adaptativo vs. Maladaptativo).

En síntesis, en el modelo de Dweck y Leggett las teorías implícitas de los individuos juegan un papel fundamental en la adopción de metas, las que a su vez conducen a diferentes patrones motivacionales. Estas teorías implícitas de los sujetos median su percepción sobre sus experiencias de aprendizaje, así como también, se relacionan con diferentes formas en que los sujetos pueden aprender. Por tanto, de acuerdo a nuestra hipótesis central de investigación las auto-teorías de los sujetos consideradas en el modelo de Dweck son variables relevantes para evaluar nuestra hipótesis de investigación. A su vez, el modelo de Dweck y Leggett es una de las perspectivas teóricas útil en dicho proceso.

2.5. Modelos de Rasch y estudio del proceso de respuesta a los SET.

Existe otro aspecto clave que es importante considerar para aproximarse al problema de investigación abordado. Este aspecto está relacionado con encontrar una aproximación que permita modelar el proceso de respuesta a nivel individual. También, implica lograr considerar las variables que puedan mediar este proceso de respuesta.

Esto no es posible con los métodos corrientemente usados para analizar los

SET. En estudios de puntajes estudiantiles, en general, se consideran modelos que tienen como variable respuesta la media grupal de puntajes estudiantiles. En este caso la unidad de análisis corresponde a grupos de estudiantes. Por otra parte, estos modelos tampoco consideran la distribución conjunta del patrón de respuesta a los ítems.

Por ejemplo, si se aplica un modelo de regresión lineal múltiple al análisis de estos datos, la variable respuesta sería el puntaje promedio asignado a cada docente en un ítem o un conjunto de ítems por un grupo de estudiantes. En este tipo de modelos, se pueden considerar variables que corresponden a características de los estudiantes (a nivel grupal), docentes y el contexto (Chang, 2000). Además, estos modelos de regresión no consideran las posibles correlaciones que pueden existir entre las respuestas a los ítems. Por lo tanto, el análisis del proceso de respuesta de cada estudiante queda totalmente fuera de alcance con estas aproximaciones.

Una aproximación distinta la ofrece una familia de modelos probabilísticos de los procesos de respuesta a ítems, denominada Modelos de Rasch. Estos modelos relacionan la probabilidad de respuesta de cada sujeto a cada ítem con una variable latente, que es una variable no directamente observable (Eve-ritt, 1984). En los Modelos de Rasch unidimensionales, los efectos que actúan sobre el proceso de respuesta a los ítems son aditivos y se pueden expresar en términos de una variable latente unidimensional.

De esta forma, para cada elemento involucrado se obtiene una medida en un mismo marco de referencia. La medida que se obtiene para los diferentes elementos tiene una unidad constante y constituye una escala de intervalo, denominada escala logit (Engelhard Jr, 2013). Debido a que es una escala de intervalo es posible transformarla en otra haciendo un cambio de origen y un cambio en la unidad (Andrich, 1988).

Burry y Shaw (1988) aplicaron un modelo de Rasch (“Rating Scale Model”) para estudiar la validez de medidas de efectividad en la enseñanza. Una limitación de la aproximación que siguieron estos autores, con relación a estudiar el proceso de respuesta a los SET, es que el modelo aplicado no permite considerar el efecto de otras variables sobre el proceso de respuesta. Otra limitación importante, aún más fundamental, es que el modelo que emplearon no considera el efecto de los estudiantes (evaluadores) en el proceso de respuesta.

Una mejor aproximación a la modelización del proceso de respuesta la ofrece el modelo de Rasch de Múltiples Facetas (modelo MFRM). Este modelo pro-

puesto por [Linacre \(1989a, 1989b, 1994\)](#) permitiría estimar simultáneamente: (a) parámetros que describen la efectividad del docente o curso con respecto a una dimensión de efectividad; (b) la dificultad de que el evaluador (el estudiante) asigne una categoría de efectividad K relativa a la categoría $K-1$, para cada una de las categorías consideradas; (c) la severidad de cada evaluador; (d) otros parámetros que permiten considerar otras variables relacionadas con el evaluador, el docente evaluado, u otras variables que puedan estar relacionadas con el proceso de respuesta ([Bond y Fox, 2015](#); [Engelhard Jr, 2013](#)).

El modelo MFRM asume que cada observación ordinal es el resultado de la interacción de diferentes elementos. Por ejemplo, en el problema de investigación considerado, los elementos básicos son: docentes, estudiantes e ítems. En el modelo MFRM esta interacción se modela considerando que cada elemento opera en forma independiente. Además, se considera que las medidas de los elementos se combinan en forma aditiva en la variable latente ([Linacre, 2014](#)).

El modelo MFRM es además un modelo de medidas invariantes ([Andrich, 1988](#); [Engelhard Jr, 2013](#)). Esto significa en nuestro problema de investigación, por ejemplo, que las medidas de los ítems no dependen de la muestra específica de estudiantes y docentes considerados. En forma similar, significa para las medidas de los docentes, que éstas no dependen de la muestra específica de ítems y estudiantes considerados.

La propiedad de invarianza de las mediciones y el hecho que los modelos de Rasch no se basan en su formulación en asumir un muestreo aleatorio o randomización permite obtener estimaciones insesgadas de las medidas a partir de muestras no representativas ([Embretson y Reise, 2000](#)).

Otras consecuencias de las propiedades de los modelos de Rasch es que la determinación de las características y el tamaño de las muestras de calibración se determinan en forma diferente a las técnicas usuales de muestreo. Por ejemplo, en general no existen formulas para determinar el tamaño de la muestra. En su lugar, existen orientaciones generales y criterios. Estos criterios en general son elaborados utilizando simulaciones mediante métodos de Monte Carlo ([French y Dodd, 1998](#); [Manly, 2006](#)).

Además, al igual que otros modelos de Rasch, el modelo MFRM, se basa en la independencia condicional de las observaciones. Esto implica, que la probabilidad condicional del patrón de respuesta a un conjunto de ítems, se considera igual al producto de las probabilidades condicionales de cada una de las observaciones que forman parte del patrón de respuesta.

La independencia condicional es fundamental en la aplicación de modelos de Rasch. Es la base para la estimación de los parámetros por métodos de máxima verosimilitud y para las pruebas de hipótesis estadísticas (Baker y Kim, 2004). Así, aplicando la independencia condicional se puede determinar la probabilidad condicional del patrón de respuesta a un conjunto de ítems, como el producto de las probabilidades condicionales de las observaciones correspondientes a cada ítem, por ejemplo:

$$\Pr(X_{nj1k}, X_{nj2k} | \theta_n, \alpha_j, \beta_i, \tau_k) = \Pr(X_{nj1k}, | \theta_n, \alpha_j, \beta_1, \tau_k) \cdot \Pr(X_{nj2k}, | \theta_n, \alpha_j, \beta_2, \tau_k) \quad (2.1)$$

La 2.1 expresa la probabilidad condicional de responder a dos ítems ordinales $i = 1$ e $i = 2$, a un nivel k , cuando un estudiante j evalúa a un docente n . Este tipo de ecuaciones se utilizan en los algoritmos de máxima verosimilitud, donde se definen las funciones que dan la probabilidad del patrón de respuesta en función de las medidas a estimar. Los valores estimados de las medidas corresponden a valores que maximizan la probabilidad del patrón de respuesta.

Una diferencia importante con otros modelos es que este modelo refiere al proceso de respuesta que involucra una serie de elementos que participan. Adicionalmente, una visión común en la aplicación de estos modelos, denominada interpretación estocástica del sujeto, interpreta que la distribución de probabilidades dada por el modelo aplica a los sujetos individuales (Ellis y Van den Wollenberg, 1993; Holland, 1990; Lord y cols., 1968; Samejima, 1983). Esto contrasta con técnicas de inferencia estadística usadas habitualmente. De esta forma, en las técnicas de inferencia estadística usadas más habitualmente, el modelo estadístico implícito se basa en la suposición de cierto tipo de muestreo aleatorio a partir de una población, o se basan en la randomización de tratamientos a grupos de sujetos.

Por el contrario, la aplicación de modelos de Rasch para obtener medidas invariantes, no se basa en la obtención de muestras aleatorias, ni en la randomización. Adicionalmente, no se busca estimar un parámetro poblacional a partir de una muestra. Por el contrario, el modelo se utiliza para medir los efectos que tienen cada uno de los elementos involucrados en el proceso de respuesta. En el caso del modelo MFRM, se busca estimar el efecto de cada docente, ítem y estudiante, en la probabilidad de cada observación.

Otra diferencia importante, tiene que ver con el enfoque de modelización

seguido, cuando se trabaja con modelos de Rasch y medidas invariantes (Bond y Fox, 2015; Engelhard Jr, 2013). En otras aproximaciones se estudia el ajuste de un modelo a los datos considerados. Si el ajuste no es bueno, se modifica el modelo para intentar resumir o explicar los datos. Por el contrario, con modelos de Rasch, el enfoque es estudiar en qué grado los datos cumplen con cierto ideal para obtener medidas y construir una escala. De alguna manera, la aproximación seguida se asemeja más a un control de calidad. No se busca un ajuste perfecto, pues, los modelos establecen un ideal. Lo que se busca es ver si los datos tienen un ajuste suficiente como para que se puedan obtener medidas útiles.



Capítulo 3

Metodología.

3.1. Muestra.

Los datos que se utilizaron en esta investigación son datos secundarios que provienen del Sistema de Evaluación de la Enseñanza y Proceso de Aprendizaje de Facultad de Medicina. El abordaje de problemas vinculados a evaluaciones institucionales de la enseñanza corrientemente sigue la aproximación propuesta, basada en la utilización de datos secundarios provenientes de sistemas de evaluación institucional (Bianchini, Lissoni, y Pezzoni, 2013; E. H. Cohen, 2005; Marsh y Hocevar, 1991; Obiekwe, 1999). La aproximación seguida no implicó dedicación especial de tiempo de los estudiantes para aportar datos a esta investigación. Tampoco la realización de esta investigación alteró, en forma alguna, las actividades que habitualmente realizan los estudiantes en los cursos.

3.1.1. Funcionamiento del Sistema de Evaluación y datos recabados a partir del año 2011.

Desde el 2011 se aplican en Facultad de Medicina escalas de evaluación de la enseñanza y proceso de aprendizaje. En cada semestre participan aproximadamente 5000 estudiantes en las evaluaciones de la enseñanza y proceso de aprendizaje. Desde el 2011 se han aplicado escalas de evaluación correspondientes a aproximadamente 160 evaluaciones de cursos y unas 2500 evaluaciones de docentes. Se debe tomar en cuenta, que el número de cursos y docentes distintos que participan es menor, pues, algunos de los cursos y docentes han sido evaluados varias veces.

A partir de los datos recabados se realizan reportes escritos que se envían a los docentes y responsables de cursos. Los reportes de cursos además se publican en la WEB de Facultad de Medicina en un período de 6 a 12 meses posterior a que se enviaron los reportes de evaluación a los docentes responsables y participantes. Los reportes de docentes se envían vía mail y también se publican en la WEB de Facultad de Medicina. Esto último, se hace al menos dos años después que se recabaron los datos y fue enviado un reporte escrito de evaluación al docente. Esta práctica de hacer pública esta información y rendir cuentas a la sociedad es consistente con políticas de responsabilidad social universitaria (Quezada, 2011; Vallaey, 2009).

Por otra parte, para recabar los datos, se aplican las escalas mediante cuestionarios WEB a los que cada estudiante accede mediante una clave de acceso. La participación de los estudiantes no es obligatoria y sus respuestas son tratadas en forma confidencial. Además, en los cuestionarios no se registra el nombre del estudiante. Tampoco se incluyen datos personales sensibles.

Existen tres protocolos básicos de aplicación. El primero, se aplica en evaluaciones de fin de semestre. En este protocolo las evaluaciones se realizan en las últimas dos semanas de un curso semestral, previo a la evaluación final del curso.

El segundo protocolo, se aplica en evaluaciones durante el semestre, para actividades de enseñanza con menor duración. En este caso, las evaluaciones se realizan en un período de 1 a 2 semanas, cerca del final de la actividad de enseñanza y previo a la evaluación final.

El tercer protocolo, se aplica para evaluaciones de clases, generalmente teóricas, cuando muchos docentes participan de la actividad de enseñanza que comprende entre 20 y 30 sesiones aproximadamente, pero, cada docente imparte alrededor de 2 a 3 clases. En este caso, las evaluaciones de las clases dadas por un docente se inician luego de la última clase dada por el docente, y se finalizan antes de la evaluación final de dicha actividad de enseñanza.

En todos los protocolos de evaluación seguidos, las evaluaciones se realizan sobre el final de la actividad de enseñanza evaluada y antes de la evaluación final de dicha actividad.

Además, los docentes responsables y participantes en diferentes actividades de enseñanza son contactados e informados vía mail sobre las actividades de evaluación institucional de la enseñanza. Las evaluaciones docentes se realizan solamente cuando los docentes involucrados voluntariamente quieren participar

de las mismas.

Adicionalmente, se informa sobre los contenidos de las evaluaciones por escrito y vía mail, al momento de consultar sobre la voluntad de participar. También, se envía información detallada sobre las actividades de investigación que se anticipan realizar en el futuro, y se consulta a los docentes sobre su consentimiento de que los datos sean utilizados en dichas investigaciones. Esto se hace a través de los responsables de cada actividad de enseñanza. Los responsables indican su consentimiento vía mail y/o en forma oral. Además, se ofrece la posibilidad de informar a cada docente participante en una actividad presencial informativa.

Asimismo, previo a que los estudiantes decidan voluntariamente participar de las evaluaciones, se les envía información sobre las evaluaciones que pueden completar sobre las actividades de enseñanza que están cursando. Antes de completar cada cuestionario WEB aparece un texto informativo y luego un consentimiento informado (véase Apéndice 1).

El consentimiento informado WEB permite que los estudiantes indiquen su consentimiento o no, a que los datos sean utilizados con fines de investigación. En dicho consentimiento, entre otras informaciones, se explica en qué consisten los datos que serán recabados, y cómo estos se utilizarán en actividades de evaluación educativa e investigación.

Otro aspecto importante a considerar, es la tasa de respuesta a las evaluaciones. En general oscila entre un 65 % y un 95 % de los estudiantes que cursan la actividad de enseñanza evaluada, con un promedio de aproximadamente un 75 %. De los estudiantes que realizan las evaluaciones, aproximadamente el 99 % consiente en que los datos sean utilizados también para actividades de investigación.

3.1.2. Criterios de selección de la muestra de datos.

Para estimar el tamaño de la muestra de datos de calibración, para el modelo MFRM, no existe una ecuación que permita calcular su valor (véase sección 2.5). No obstante, existen orientaciones generales que permiten estimar un número aproximado de observaciones de cada elemento considerado (docentes, estudiantes e ítems). También, existen criterios que describen como deben conectarse los elementos entre sí. Si se cumple aproximadamente con dichas orientaciones, se espera que las estimaciones tengan una razonable precisión,

y que las pruebas de hipótesis consideradas tengan una potencia aceptable.

Linacre (2014) propone como orientación general que es necesario tener 30 observaciones por cada elemento considerado y 10 observaciones por categoría de puntajes. Además, plantea que es necesario que exista suficiente solapamiento aleatorio entre los elementos.

Estos criterios fueron aplicados para estimar el número de docentes y estudiantes que es necesario considerar. También, se aplicaron estos criterios para estimar cuantos semestres consecutivos sería necesario considerar.

Para estos cálculos, fue necesario considerar como es la conexión habitual en los datos en el Sistema de Evaluación. Esto está relacionado con cuantas evaluaciones hace cada estudiante, cuantos docentes suele evaluar, y cuáles de las escalas responde. También, fue necesario tomar en cuenta las tasas de respuesta habituales y las tasas de aceptación del consentimiento informado.

En relación a las escalas consideradas, los cálculos se hicieron considerando las escalas utilizadas en el Sistema de Evaluación, que se decidieron incluir en la presente investigación. Adicionalmente, los cálculos fueron realizados para las actividades de enseñanza de los primeros 3 años de la carrera. Las escalas que se seleccionaron son las de uso más frecuente y extendido en el sistema de evaluación, en los primeros tres años de la carrera.

Con esta información, el problema se transformó en un problema de conteo. Para ello se escribió un programa en lengua R (R Development Core Team, 2008), escrito por el autor de esta investigación, que permitió calcular una estimación del número de estudiantes requeridos, docentes, y el número de semestres consecutivos que sería necesario considerar.

A partir de los cálculos realizados, se obtuvo que sería necesario considerar evaluaciones disponibles de estudiantes a docentes al menos durante 2 semestres. Considerando el tiempo disponible para realizar la presente investigación, se decidió considerar datos recabados correspondientes a las respuestas de cada estudiantes durante 2 semestres consecutivos, en el año 2015.

3.1.3. Muestra de datos de calibración.

La muestra de datos obtenida correspondió a 3243 estudiantes que evaluaron en conjunto 98 docentes de 10 actividades de enseñanza distintas que pertenecen a siete disciplinas diferentes. Estos datos corresponden a estudiantes que respondieron las escalas consideradas, durante dos semestres consecutivos,

en los primeros tres años de la carrera de medicina durante el año 2015.

3.2. Instrumentos.

Los instrumentos considerados en esta investigación forman parte de las escalas utilizadas en el Sistema de Evaluación de la Enseñanza y Proceso de Aprendizaje. En particular se consideraron las escalas más frecuentemente aplicadas en los primeros tres años de la carrera de médico. Estas escalas fueron desarrolladas por el responsable del Sistema de Evaluación Institucional de la Enseñanza y Proceso de Aprendizaje, que es también, el autor de la presente investigación. Las escalas que se consideran pueden ser clasificadas en dos grupos.

El primer grupo, corresponde a escalas de evaluación de la enseñanza. Todas las escalas en este primer grupo, contienen ítems con escala Likert (Likert, 1932) con cinco categorías, que indican diferentes niveles de acuerdo con respecto a una proposición o afirmación. Las categorías consideradas y el puntaje asignado (indicado entre paréntesis) son: “fuertemente en desacuerdo” (1); “en desacuerdo” (2); “ni de acuerdo ni en desacuerdo” (3); “de acuerdo” (4); y “fuertemente de acuerdo” (5). Por otro lado, este primer grupo de escalas corresponde a la evaluación de factores generales de efectividad (o calidad) en la enseñanza. Este grupo de escalas es el más ampliamente usado.

Adicionalmente, este grupo de escalas fue desarrollado siguiendo procedimientos requeridos para obtener instrumentos con una adecuada validez de contenido y validez aparente (Wainer y Braun, 2013). Los contenidos de estas escalas fueron elaborados a partir de la consideración de una exhaustiva revisión de trabajos e instrumentos. Estas escalas fueron elaboradas para medir dimensiones efectividad o calidad de la enseñanza relacionadas con las dimensiones medidas por “Endeavor Instrument” (Frey y cols., 1975) y “Students’ Evaluation of Educational Quality” o SEEQ (Marsh, 1987). Por tanto, los principales contenidos de las escalas consideradas en esta investigación fueron derivadas principalmente de dichos instrumentos.

Además, para la elaboración de la actual versión de las escalas consideradas, se tomaron en cuenta comentarios de docentes y estudiantes. También, como parte del análisis conceptual y lógico de los contenidos de las escalas, y de elaboración de los enunciados de los ítems, se analizó las características y el contexto particular de las actividades de la enseñanza en Facultad de Medicina.

El segundo grupo de escalas, que son utilizadas para evaluación del proceso de aprendizaje, se elaboraron en forma similar a las escalas antes mencionadas. Las teorías, variables e instrumentos considerados en la elaboración de este último grupo de escalas, al igual que algunas de las referencias consideradas (R. Ames y Lau, 1979; Bandura, 2006; Black y Deci, 2000; Deci, Koestner, y Ryan, 2001; Dweck, 2000; Schunk y cols., 2014) se resumen en la Tabla 2.1 de la sección 2.4.1.

Este segundo grupo, contiene ítems con dos tipos de escala. Al igual que en el grupo anterior de escalas, se utilizan también ítems con la escala Likert, mencionada antes, con cinco categorías. Pero, también se utilizan con bastante frecuencia la siguiente escala: “nada verdadero” (1); 2, 3, 4, 5, 6 (diferentes niveles de algo verdadero) y “muy verdadero” (7). Este segundo tipo de escala de ítems se emplearon en las escalas derivadas de instrumentos basados en la teoría de la autodeterminación (Deci y Ryan, 2008; R. M. Ryan y Deci, 2000b).

3.3. Procedimientos.

3.3.1. Pre-procesamiento de los datos y anonimización.

Previo a la etapa de análisis de los datos se evaluó su consistencia. Luego, los datos fueron anonimizados. La anonimización de los datos se realizó utilizando un algoritmo de anonimización escrito en lenguaje R (R Development Core Team, 2008). Parte del código se adjunta en el Apéndice 2. El algoritmo desarrollado no es un algoritmo de encriptación o mera recodificación. Luego de aplicado, no es posible saber a qué estudiante o docente corresponde cada registro. No obstante, en los datos anonimizados a cada estudiante y docente le corresponde un código de identificación único.

Además, se re-codificaron los códigos de identificación de las actividades de enseñanza, lo que complementa el efecto del algoritmo anterior. Adicionalmente, en relación a las actividades de enseñanza se definieron dos variables auxiliares. La primera, describe la modalidad de enseñanza, clasificada en: Teórica, Taller, Práctica, Discusión Grupal, Seminario y Otra. Los nombres de estas categorías reflejan los términos que se utilizan en Facultad de Medicina para denominar modalidades de actividades de enseñanza. La segunda, clasifica la actividad de enseñanza, de acuerdo a sus contenidos, en las siguientes disciplinas: Anatomía, Bioestadística, Biofísica, Fisiología, Histología, Inmunología y

Metodología Científica.

3.3.2. Análisis de Datos.

Aproximación Analítica.

Para evaluar la hipótesis de investigación la principal aproximación fue la aplicación de modelos de Múltiples Facetas de Rasch (MFRM). La estimación de las medidas de los elementos, el diagnóstico de ajuste entre los datos y el modelo, el análisis de invarianza de las medidas, el análisis de sesgos y estructura de las escalas de los ítems, se llevó a cabo utilizando el programa Facets versión 3.71.4 (Linacre, 2015) y R version 3.4.3 (R Development Core Team, 2008). La aproximación general adoptada siguió las perspectivas de Bond y Fox (2015); Eckes (2009); Engelhard Jr (2013); Linacre (2010, 2014). Para la presentación de los resultados se siguieron además las orientaciones recomendadas para la aplicación de modelos de Rasch dadas por R. Smith, Linacre, y Smith (2003).

Estadísticos básicos considerados y su interpretación.

MAE.

MAE (error absoluto medio) es la diferencia absoluta promedio entre los valores observados y predichos por el modelo (Willmott y Matsuura, 2005). Por ejemplo, un MAE igual a 0,40 corresponde a un 10 % del rango de variación de la variable respuesta, cuando los ítems considerados tienen puntajes observados entre 1 y 5. Así, en este caso su rango es de 4, y por consiguiente, si el valor de MAE es de 0,40, representa un 10 % del rango de variación del puntaje. Adicionalmente, los puntajes observados diferentes, difieren en por lo menos 1 punto, por lo que para un MAE de 0,40 el error absoluto promedio sería menor que la mitad de la diferencia mínima entre dos puntajes distintos. Valores bajos de MAE describen un buen ajuste entre los datos y el modelo. El MAE puede normalizarse respecto de la media (o el rango) de la variable respuesta y expresarse como porcentaje de la media (o respecto al rango), y en tal caso, el valor normalizado se denomina error absoluto medio normalizado.

MBE.

MBE (sesgo promedio de los errores) es el promedio de los residuos. MBE describe la diferencia entre el promedio de los valores observados de la variable respuesta y el promedio de los valores predichos por el modelo. Por tanto, MBE permite describir si en promedio los puntajes predichos por el modelo son mayores o menores que los valores observados (Willmott y Matsuura, 2005). El MBE puede normalizarse respecto de la media de la variable respuesta (o respecto al rango) y expresarse como porcentaje de la media (o el rango), y en tal caso, el valor normalizado se denomina sesgo promedio de errores normalizado o NMBE (Reddy y cols., 1997). Valores bajos de MBE describen un buen ajuste entre los datos y el modelo. No existe un valor universalmente aceptado como un punto de corte para definir un ajuste pobre o aceptable. En aplicaciones del MBE con modelos de regresión múltiple, vinculados a problemas de predicción, una convención que se aplica en algunos casos es considerar como aceptable un valor de NMBE menor a 15% (Reddy, 2011; Reddy y cols., 1997).

RMSE.

RMSE (raíz del error cuadrático medio) es la raíz cuadrada de la media de los cuadrados de los errores. RMSE para cada faceta se determina como la raíz cuadrada de la media de los cuadrados de los errores estándar de las medidas de cada elemento de la faceta (B. D. Wright y Stone, 1999). En este caso se expresan en logits. Cuanto menor es el valor de RMSE de una faceta, mejor es el ajuste de los elementos de dicha faceta. En particular, un $RMSE < 0,10$ ha sido utilizado como criterio para definir una precisión aceptable de la estimación de los parámetros de los ítems (De Ayala, 2013).

RMSE para el modelo corresponde a la raíz cuadrada de la media de los cuadrados de los residuos. La definición específica empleada y su interpretación es discutida en detalle por Willmott y Matsuura (2005). Los residuos son la diferencia entre los puntajes observados y los predichos por el modelo. Un ejemplo de aplicación en modelos de Rasch, pero para evaluar el error de estimación de parámetros del modelo, puede verse en (Babcock y Albano, 2012). RMSE para el modelo se puede interpretar como el desvío estándar de los valores observados de la variable respuesta respecto a los predichos por el modelo. Cuanto más cercano a cero es el valor obtenido, mayor ajuste existe

entre los datos y el modelo. El valor de RMSE para el modelo depende del valor de la variable respuesta considerada y tiene las mismas unidades que esta. No existe un punto de corte universal que defina un ajuste aceptable o bueno.

CV(RMSE).

CV(RMSE) es el coeficiente de variación de RMSE para el modelo, esto es, el valor normalizado de RMSE para el modelo respecto a la media de la variable respuesta. Por ejemplo, un valor de 10 % significa que la variabilidad no explicada por el modelo es un 10 % de la media. Para modelos de regresión múltiple, en aplicaciones orientadas a la predicción, se han definido algunos criterios generales para interpretar este estadístico que podrían aplicarse para definir criterios de bondad de ajuste global (Draper y Smith, 2014; Reddy, 2011; Reddy y cols., 1997; Reynolds y Fels, 1988).

Para interpretar este estadístico aquí se considerarán como orientación general puntos de corte análogos a los definidos por Reddy y cols. (1997). Así, se considerarán las siguientes categorías de ajuste de acuerdo al valor de CV(RMSE): ajuste pobre (CV(RMSE) mayor o igual a 20 %); ajuste aceptable (CV(RMSE) mayor o igual que 10 % y menor que 20 %); ajuste bueno (CV(RMSE) mayor que 5 % y menor que 10 %); y ajuste excelente (CV(RMSE) igual o menor que 5 %).

Una normalización alternativa de RMSE puede realizarse respecto al rango de la variable respuesta. En este caso el valor normalizado se denomina valor normalizado de RMSE (NRMSE). Los valores de NRMSE suelen ser próximos a los valores de CV(RMSE). No obstante, el valor de NRMSE expresa el valor de RMSE tomando como unidad el rango de la variable respuesta en lugar de la media.

Proporción de residuos estandarizados absolutos extremos (PRE_x).

PRE_x es la proporción observada de los valores absolutos de los residuos estandarizados mayores a cierto valor x . En particular, si el ajuste entre los datos y el modelo es perfecto se espera que la distribución de los residuos estandarizados sea normal estándar (B. D. Wright y Stone, 1999). Por tanto, si el modelo tiene un buen ajuste, se espera que PRE_3 sea próxima al valor 0.003, mientras que PRE_2 sea próxima al valor 0.046, que corresponden a las

proporciones esperadas si el ajuste entre los datos y el modelo es perfecto, para valores absolutos de los residuos estandarizados 3 y 2, respectivamente.

Proporción de la varianza del puntaje observado explicada por el modelo (PVE).

PVE es la proporción de la varianza de los puntajes observados explicada por el modelo, por tanto, varía entre 0 y 1. Si $PVNE$ es la proporción de la varianza del puntaje observado no explicada por el modelo, entonces $PVE + PVNE = 1$. El valor de PVE es suministrado por el programa Facets. El cálculo de PVE es igual al cálculo del coeficiente de determinación o R^2 para modelos de regresión lineal múltiple.

PVE podría ser considerado como un análogo para variables politómicas al *pseudo* - R^2 propuesto por Efron (1978) para variables dicotómicas. En este caso el valor esperado de acuerdo al modelo para la variable politómica ordinal es calculada como se muestra en B. D. Wright y Masters (1990).

Cuanto mayor es el valor de PVE mayor es la proporción de la varianza de los puntajes observados explicada por el modelo. PVE es relevante si nos interesa entender cuán bueno es el modelo explicando el patrón de respuesta observado. La interpretación de PVE es similar a la interpretación de R^2 . No obstante, debido a que el modelo es probabilístico, en ciertas condiciones se espera que el valor de PVE sea bajo, a pesar de que el modelo prediga correctamente las probabilidades de respuesta a cada ítem (Linacre, 2006).

La interpretación de la magnitud del tamaño del efecto que significan diferentes valores de PVE para un modelo de Rasch se puede realizar siguiendo la orientación general (Linacre, 2005): 2 % (tamaño del efecto pequeño); 13 % (tamaño de efecto medio); y, 26 % (tamaño de efecto grande). Esta orientación general es similar a otra sugerida por J. Cohen (1992) para modelos de regresión lineal, para el índice de tamaño de efecto $f_2 = R_2/(1 - R_2)$: 2 % (tamaño del efecto pequeño); 15 % (tamaño de efecto medio); y, 35 % (tamaño de efecto grande). El criterio sugerido por J. Cohen (1992) expresado en términos de R_2 es: 2 % (tamaño del efecto pequeño); 13 % (tamaño de efecto medio); y, 26 % (tamaño de efecto grande). Observar que los valores sugeridos por (Linacre, 2005) coinciden con los valores sugeridos por J. Cohen (1992). La diferencia entre ambos criterios es que PVE es un *pseudo* - R_2 definido para medidas de Rasch que generaliza la definición dada por Efron (1978), a variables categóricas.

cas ordinales.

Infit y Outfit.

Infit y Outfit (Infit MnSq y Outfit MnSq en el programa Facets) son estadísticos de ajuste entre los datos y el modelo que pueden ser calculados a nivel de los elementos de cada faceta, o bien, pueden calcularse para todo el modelo como una medida de ajuste global entre los datos y el modelo. Los valores de Infit y Outfit son suministrados por defecto por el programa Facets para cada elemento de las facetas consideradas. Para el cálculo de Infit y Outfit, como medidas de ajuste global del modelo, es necesario especificar el modelo con una faceta “dummy” con un único elemento, presente en todas las observaciones (de Jong y Linacre, 1993).

Infit y Outfit proveen información sobre las discrepancias que ocurren en las respuestas observadas y las esperadas de acuerdo al modelo. Infit le da más peso a las discrepancias que ocurren en las respuestas cerca de los parámetros estimados. Por el contrario, Outfit, le da más peso a las discrepancias observadas en las respuesta lejos de los parámetros estimados. Infit es un estadístico de ajuste ponderado por información, basado en el promedio de los cuadrados de los residuos estandarizados entre lo que es observado y lo que es esperado de acuerdo al modelo. Outfit es un estadístico similar al Infit, pero, no es ponderado. Infit y Outfit se calculan promediando estadísticos con distribución “chi-cuadrado”, lo que da como resultado estadístico de cuadrados medios. El detalle de los cálculos involucrados puede verse en (B. D. Wright y Masters, 1990).

Los valores de Infit y Outfit muestran el tamaño de la aleatoriedad de los datos. Si el ajuste entre los datos y el modelo es perfecto el valor esperado es 1. Si el valor es mayor que 1 la aleatoriedad observada en los datos es mayor a la esperada y sugiere un bajo ajuste (*underfit*). Por el contrario, si el valor es menor que 1, la aleatoriedad en los datos es menor que la esperada y sugiere que existe un sobre ajuste (*overfit*). De acuerdo con Linacre (2014) valores entre 0,5 y 1,5 representan un ajuste satisfactorio. Un criterio más exigente es el propuesto por R. M. Smith y cols. (1995), donde los puntos de corte para el Infit y Outfit son respectivamente $1 \pm \frac{2}{\sqrt{N}}$ y $1 \pm \frac{6}{\sqrt{N}}$, donde N es el número de observaciones consideradas.

Correlación punto - medida (r_{pm}).

La correlación punto - medida (r_{pm}) es un estadístico de Rasch análogo a la correlación punto - biserial (r_{pbis}). r_{pm} permite evaluar si la medida de un elemento es consistente ($r_{pm} > 0$) o contradice ($r_{pm} < 0$) el significado general de un test. Adicionalmente, r_{pm} para los ítems de un test puede ser usado para detectar errores de codificación. r_{pm} tiene una interpretación similar al r_{pbis} , pero, presenta diversas ventajas (Linacre, 2006).

Medidas de Invarianza.

En teoría los parámetros del modelo (medidas de los elementos de las facetas) son invariantes. No obstante, la invarianza efectivamente observada depende del ajuste entre los datos y el modelo. Debido a esto, las evidencias de invarianza informan sobre el nivel de ajuste entre los datos y el modelo. Si las medidas obtenidas son efectivamente invariantes, se espera una correlación lineal elevada entre las medias obtenidas a partir de diferentes submuestras. En particular, se espera un coeficiente de correlación lineal elevado entre dos sub-muestras aleatorias complementarias, así como, coeficientes de correlación lineal elevados entre cada una de las sub-muestras y la muestra completa (Rupp y Zumbo, 2004). Un coeficiente de correlación de Pearson entre dos sub-muestras mayor a 0,95 puede interpretarse como un buen ajuste, mientras que uno mayor o igual a 0,98 puede considerarse un ajuste excelente o notable (Muñiz, 1997). La correlación lineal entre dos sub-muestras aleatorias complementarias es designada con $r(1,2)$. Las correlaciones entre las sub-muestras y la muestra completa son representadas con $r(1,C)$ y $r(2,C)$.

Confiabilidad de separación (SR), Razón de Separación (G), Estratos (H) y estadísticos de chi-cuadrado.

Confiabilidad de separación (SR). SR denota la confiabilidad de separación o confiabilidad de Rasch. SR es la razón entre la varianza verdadera y la varianza observada de las medidas de los elementos de una faceta. SR muestra cuán reproducible es el ordenamiento entre las medidas de los elementos de una faceta. Su valor varía entre 0 y 1. Cuanto mayor es el valor de SR mayor en la confiabilidad de la medida.

En particular, SR para los elementos que son objeto de evaluación dan el equivalente de la confiabilidad del test en la teoría clásica de los test. SR es

el análogo al coeficiente alpha de Cronbach y tiene una interpretación similar (Linacre, 2006; B. D. Wright y Stone, 1999). No obstante, SR tiene diversas ventajas respecto a medidas de confiabilidad obtenidas en teoría clásica de los test (B. D. Wright y Stone, 1999).

Como orientación general, se puede considerar para un test: un valor de confiabilidad mayor a 0,7 como aceptable; entre 0,80 y 0,90 como buena; y un valor de confiabilidad mayor a 0,9 como elevada (Bernstein y Nunnally, 1994). En esta investigación se utilizarán estos intervalos de valores, como criterio general y primera aproximación, para interpretar los valores de SR de las facetas.

Razón de Separación (G). G es la razón entre el desvío estándar ajustado ($SD_A = SD - MSE$) y la raíz del error cuadrático medio de la medida ($RMSE$). Este índice expresa en escala de razón la confiabilidad de la medida tomando como unidad el valor de $RMSE$ (B. D. Wright y Stone, 1999). Su relación con SR está dada por: $G = \sqrt{\frac{SR}{1-SR}}$.

Índice de Estratos (H). H es el número de estratos distintos medibles que pueden obtenerse con los datos (Bond y Fox, 2015). H se puede calcular con la formula: $H = \frac{4G+1}{3}$.

Chi-cuadrado efectos fijos (iguales). El test de chi-cuadrado efectos fijos (iguales) permite contrastar la hipótesis que los elementos de una faceta tienen la misma medida, atribuyendo las diferencias observadas a los errores de estimación (Linacre, 2006).

Chi-cuadrado efectos aleatorios (normales). El test de chi-cuadrado efectos aleatorios normales permite contrastar la hipótesis de que los elementos de una faceta tienen medidas que pueden considerarse una muestra aleatoria de una distribución normal (Linacre, 2006).

Interacción entre elementos de las facetas.

PVE_{INTER} . PVE_{INTER} es la proporción de la varianza de los puntajes observados explicada por el término de interacción en el modelo. Si PVE es la proporción de la varianza explicada por el modelo que contiene solo efectos primarios, $PVE + PVE_{INTER}$ es igual a la proporción de la varianza explicada por el modelo con la interacción o PVE_{MCI} . El valor de PVE_{INTER} nos permite valorar la importancia que tiene el término de interacción en la explicación de los puntajes observados.

Chi-cuadrado efectos fijos (todos cero). El test de chi-cuadrado efectos fijos (todos cero) permite contrastar la hipótesis de que los términos de interacción tienen todos medida cero, atribuyendo las diferencias observadas a los errores de estimación (Linacre, 2006).

Estadísticos de validez de las categorizaciones.

Los estadísticos de validez de las categorizaciones son utilizados para analizar si los sujetos que responden a dichos ítems hacen un uso consistente de las categorías definidas. Por ejemplo, la categoría “muy de acuerdo” representa realmente un nivel mayor del constructo medido que la categoría “de acuerdo”; o bien, para cierto ítem, los sujetos que responden no pueden distinguir entre ambas categorías.

Promedio de medidas de las observaciones en la categoría. Los promedios de las medidas de las observaciones en las categorías deben estar ordenados en forma creciente para poder interpretar que una mayor categoría

corresponde a un mayor valor del constructo medido (Bond y Fox, 2015; Linacre y cols., 2002).

Umbral de Rasch-Aldrich. Otro indicador de la validez de las categorizaciones se basa en los umbrales de Rasch-Aldrich. Se considera que umbrales de Rasch-Aldrich ordenados en forma monótona con la categoría de puntaje indican una mayor calidad de la escala (Bond y Fox, 2015; Eckes, 2009; Linacre y cols., 2002). Adicionalmente, se considera como indicador de calidad de la escala de puntajes que la separación entre los umbrales sea al menos 1 (para escalas de puntajes con 5 categorías) y menor que 5 logits (Linacre y cols., 2002).

Outfit de la categoría. El valor esperado de Outfit para una categoría es 1. Valores de Outfit de la categoría mucho mayores que uno indican la presencia de observaciones inesperadas en esta categoría. Valores no mayores que 2,0 se consideran valores aceptables de Outfit (Bond y Fox, 2015; Eckes, 2009; Linacre y cols., 2002).

Frecuencia de respuestas entre categorías. Distribuciones regulares de frecuencias de respuestas entre categorías, sin la presencia de categorías no observadas, se consideran un indicador de calidad de la escala de puntajes (Bond y Fox, 2015; Eckes, 2009; Linacre y cols., 2002).

3.3.3. Consideraciones Éticas.

Esta investigación se basó en el análisis de datos secundarios. Estos datos hubieran sido de todas maneras recabados para tareas de evaluación educativa, mediante la participación libre y voluntaria de los estudiantes interesados en apoyar las actividades de evaluación educativa en Facultad de Medicina.

De forma alguna se afectó con la realización de esta investigación a los estudiantes. Por un lado, este diseño permitió que no sea necesario recabar datos especialmente para llevar a cabo esta investigación. Por otro lado, los datos originales no contenían datos personales sensibles.

Adicionalmente, los datos originales no contenían información sobre el estudiante que pueda identificarlo en forma alguna, aparte de código numérico de identificación. Esto se debe a que las respuestas a estas escalas mayormente no miden atributos de quienes responden, ya que son respuestas que dependen no solo del estudiante, sino también de variables de contexto. Además, las respuestas mayoritariamente no refieren al propio estudiante, sino a sus

valoraciones subjetivas sobre la enseñanza.

También, debido a que no se trabajó con los datos originales, sino con una matriz de datos anonimizada con una aproximación rigurosa, no existe forma alguna de que se identifique a los sujetos que originalmente aportaron los datos.

Por otro lado, es importante observar que en la obtención de datos en el Sistema de Evaluación de la Enseñanza y Proceso de Aprendizaje, se informa de forma detallada de las actividades de evaluación educativa e investigación que se proyectan realizar. Además, todos los involucrados participan en forma voluntaria y debidamente informada.

En particular, cada uno de los estudiantes recibe información detallada, y se aplica un consentimiento informado WEB (véase Apéndice 1), en el cual cada estudiante puede manifestar su aceptación o no aceptación a que los datos aportados puedan ser usados en actividades de investigación.

Asimismo, el objeto de investigación en esta propuesta, es la validez de las interpretaciones y aplicaciones de las medidas obtenidas con los instrumentos utilizados en tareas de evaluación educativa.

El proyecto de investigación que incluye la investigación aquí reportada fue evaluado y aprobado por el comité de ética de investigación de Facultad de Medicina.

Capítulo 4

Hacia una teoría socio-cognitiva del proceso de respuesta a cuestionarios de evaluación estudiantil de la calidad de la enseñanza.

4.1. Presentación.

En este capítulo se proponen cuatro modelos teóricos sobre el proceso de respuesta a los SET basados en teorías cognitivas de la motivación discutidas en el capítulo 2 de esta tesis. Los modelos propuestos especifican desde cierta perspectiva teórica, la hipótesis de validez para los SET a un nivel individual. Esto es, tomando como unidad de análisis el proceso de respuesta de cada estudiante.

Bajo una hipótesis de validez de los SET, si existen diferencias en la severidad de los estudiantes, estas, deberían tener una relación inversa con el efecto facilitador de la enseñanza en el aprendizaje de cada estudiante. La hipótesis de validez de los SET constituye una especificación o restricción de la hipótesis central de esta investigación.

Cada uno de los modelos propuestos permite relacionar constructos fundamentales para el modelo o teoría en que se basa, por un lado, con el efecto facilitador de la enseñanza y la severidad del estudiante al evaluar la enseñanza,

por otro.

En la sección 4.2 se presenta el modelo atribucional del proceso de respuesta a los SET. Este modelo considera como constructo fundamental las atribuciones causales y se basa en la teoría de las atribuciones causales de Bernard Weiner discutida en la sección 2.4.3. La sección 4.3 presenta el modelo SDT del proceso de respuesta a los SET. Este modelo se basa en la teoría de la autodeterminación (véase sección 2.4.4). Luego, la sección 4.4 propone el modelo TAS fundamentado en la teoría del aprendizaje social de Bandura (véase sección 2.4.2). El último modelo propuesto es el modelo de autoteorías del proceso de respuesta a los SET, fundamentado en el modelo de Dweck y Leggett, que fue descrito en la sección 2.4.5. Finalmente, este capítulo finaliza con una discusión de las implicaciones teóricas y prácticas de los modelos propuestos.

4.2. Modelo atribucional para el proceso de respuesta a los SET.

Una predicción básica puede deducirse a partir de la hipótesis de investigación propuesta y lo discutido en la sección 2.4.3: las diferencias en el patrón de atribución a causas específicas, y en la configuración de las dimensiones causales percibidas, en relación al aprendizaje y desempeño en un curso, deberían explicar diferencias individuales en el proceso de respuesta a los SET.

Un aspecto del proceso de respuesta de los estudiantes es su severidad. La severidad esta relacionada con su tendencia a evaluar más negativamente la calidad de la enseñanza. Un corolario de la hipótesis propuesta es que el patrón de atribuciones a causas específicas, y la configuración en las dimensiones causales percibidas, serían factores de severidad. Es importante observar que la hipótesis de investigación propuesta no es direccional. Para evaluar esta hipótesis, específicamente en relación al papel de las atribuciones causales y sus dimensiones, necesitamos evaluar si explican diferencias en la severidad.

Si además se introduce una hipótesis de validez de los SET, debería existir una relación negativa entre la severidad y la calidad del proceso de aprendizaje en el contexto de un curso. Así, estudiantes que se ven más beneficiados por la enseñanza y tienen una calidad del proceso de aprendizaje mayor, se esperaría que fueran menos severos, que estudiantes con un efecto menos positivo de la

enseñanza y con una menor calidad del proceso de aprendizaje.

De esta forma, si bien se espera que existan diferencias en la evaluación realizada por cada estudiante, éstas no serían sesgos en la evaluación. Más bien, representarían diferencias en la evaluación que reflejan diferencias en el efecto de la enseñanza sobre el aprendizaje del estudiante.

Observar que la hipótesis de validez de los SET es más restrictiva que la hipótesis de investigación propuesta en esta investigación. La hipótesis principal propuesta considera además que pueden existir diferencias en el proceso de respuesta derivadas de diferencias en el procesamiento de información cognitiva. Por tanto, esta hipótesis incluye como posibilidad que existan diferencias en la evaluación realizada por los estudiantes que no necesariamente reflejen diferencias en la calidad de la enseñanza y del proceso de aprendizaje en el contexto de un curso.

A partir de algunas observaciones, la discusión teórica previa puede completarse con la propuesta de un modelo cognitivo atribucional para el proceso de respuesta a los SET. Una primera observación es que las dimensiones causales podrían tener un vínculo con las señales informativas utilizadas en el proceso de atribución (Frieze y Weiner, 1971; Meyer, 1978, 1980; Weiner, 2013). En el modelo propuesto, una primera premisa, es que la representación mental de la atribución a una causa específica se basa en parte en las dimensiones causales. Si esto fuera así, las dimensiones causales no solo serían útiles explicando las consecuencias de las atribuciones, sino también, tendrían un papel en el procesamiento de información cognitiva. La representación mental de una causa específica, al menos en parte, correspondería a cierta configuración de las dimensiones de causalidad.

Una segunda premisa del modelo propuesto, es que las configuraciones de las dimensiones causales podrían formar parte de representaciones mentales del sujeto, involucradas en la valoración de la calidad de la enseñanza. Esto es, las configuraciones causales serían parte de la información cognitiva utilizada para evaluar la calidad de la enseñanza. Los resultados obtenidos por R. Ames y Lau (1979) son compatibles con esta segunda premisa. R. Ames y Lau (1979) encontró que estudiantes que consideran de mayor importancia a causas internas como la habilidad, el esfuerzo y la intención, evalúan en forma más positiva, que estudiantes que consideran de mayor importancia causas externas como la dificultad y la suerte.

Adicionalmente, en el modelo se asume que la evaluación de la calidad de la

enseñanza se basaría en parte en la comparación entre la configuración de las dimensiones causales con una buena calidad de enseñanza y la configuración que el sujeto infiere en relación a la enseñanza de un docente en un curso. Si esta última configuración tiene una discrepancia importante con la primera, esto se relacionaría con una mayor severidad del estudiante al evaluar la calidad de la enseñanza.

En términos de la configuración de dimensiones causales, el patrón atribucional I (locus interno, inestable y controlable) podría relacionarse con una percepción de mayor calidad de la enseñanza, mientras que el patrón atribucional II (locus externo, estable e incontrolable), podría relacionarse con una percepción de menor calidad de la enseñanza. A su vez, un estudiante con un patrón atribucional I sería menos severo que un estudiante con un patrón atribucional II.

Los patrones atribucional I y II pueden corresponder a diferentes patrones de atribución a causas específicas como se ilustra en la Tabla 4.1. Por tanto, en principio sujetos con un patrón atribucional I o II podrían corresponder a diferentes combinaciones de causas específicas. Por ejemplo, un estudiante que considera como causa más importante el esfuerzo, tendría un patrón atribucional I. Pero, un estudiante que percibe que las principales causas relacionadas con su aprendizaje y desempeño son el esfuerzo, uso de estrategias y la habilidad, podría tener también un patrón atribucional I, y una configuración de las dimensiones causales, similar al ejemplo anterior.

No obstante, se debe tener presente que una causa específica puede ser percibida por sujetos distintos como diferente en locus, estabilidad y control. Por ejemplo, algunos sujetos son capaces de realizar grandes esfuerzos en forma sostenida en el tiempo. En este caso, el esfuerzo para estos sujetos podría percibirse como relativamente estable, y no inestable, como se clasifica en la Tabla 4.1. Adicionalmente, algunos sujetos con orientación a metas de aprendizaje, consideran que la habilidad se puede mejorar, mediante el esfuerzo, la búsqueda de desafíos y la persistencia. Para estos sujetos claramente la habilidad sería percibida como relativamente inestable.

Tabla 4.1: Relación entre Patrones Atribucionales I y II, valores característicos de las dimensiones causales y causas específicas más importantes para cada patrón atribucional. Para cada patrón se muestra entre paréntesis el valor característico para cada dimensión. Además se muestran las causas específicas que podrían ser más importantes para cada patrón.

| | Locus | Estabilidad | Controlabilidad |
|-----------|----------------------------------------------------------------|-----------------------------------------------------|-------------------------------------------------------|
| Patrón I | (interno) esfuerzo estrategias intención habilidad | (inestable) esfuerzo estrategias intención | (controlable) esfuerzo estrategias intención |
| Patrón II | (externo) dificultad suerte docente | (estable) dificultad habilidad docente | (incontrolable) suerte docente habilidad |

El modelo propuesto aquí se diferencia en varios aspectos del modelo propuesto por R. Ames y Lau (1979). En primer lugar, el presente modelo considera las tres dimensiones de las atribuciones causales, mientras que el modelo de R. Ames y Lau (1979) considera solo la dimensión locus. Adicionalmente, el presente modelo especifica el papel que podrían tener las dimensiones causales en la representación mental de las atribuciones causales y en el proceso de evaluación subjetiva de la calidad de la enseñanza. Por el contrario, el modelo propuesto por R. Ames y Lau (1979) no explicita exactamente como se relacionaría el procesamiento de información cognitiva vinculada a las atribuciones causales y sus dimensiones, con el procesamiento de información cognitiva relacionado con la evaluación de la calidad de la enseñanza.

4.3. Modelo SDT del proceso de respuesta a los SET.

El propósito de esta sección es describir un modelo del proceso de respuesta a los SET desde la perspectiva de la SDT. En su conjunto permite establecer un vínculo entre los principales constructos de la SDT, por un lado, y el efecto facilitador de la enseñanza y la severidad de los estudiantes, por otro. A este modelo lo denominaremos modelo SDT del proceso de respuesta a los SET. Las

Figuras 4.1, 4.3 y 4.2 resumen las principales características de este modelo.

El modelo SDT puede considerarse como una especificación, o restricción, de la hipótesis central considerada en esta investigación. Este modelo es consistente con una hipótesis de validez de los SET, expresada desde la perspectiva de la SDT. La hipótesis de validez implica que si existen diferencias en la severidad de los estudiantes al evaluar la enseñanza, está debería ser inversa al efecto facilitador de la enseñanza para el proceso de aprendizaje de cada estudiante. Esto es, para un estudiante que la enseñanza tiene un menor efecto facilitador de su aprendizaje, tendría una severidad mayor; mientras que, para un estudiante que la enseñanza tiene un efecto facilitador mayor, su severidad sería menor. De esta forma, la percepción de la enseñanza de cada estudiante primariamente reflejaría el efecto facilitador de la enseñanza para este estudiante.

Adicionalmente, el modelo SDT esta compuesto por tres componentes: el primer componte está vinculados a la mini-teoría de la integración orgánsmica (OIT) de la SDT; el segundo componente está relacionado con la mini-teoría de necesidades básicas (BPNT); y por último, el tercer componente está relacionado con la teoría de contenidos de metas (GCT).

Las características básicas del primer componente del modelo se muestran en la Figura 4.1. De acuerdo a este modelo se espera que los estudiantes motivados intrínsecamente sean menos severos que los estudiantes motivados extrínsecamente, y estos, menos severos que estudiantes amotivados. Dentro de los estudiantes motivados extrínsecamente, los tipos de regulación que reflejan un mayor grado de internalización serían menos severos que los estudiantes con un nivel menor de internalización.

Observar que los estudiantes amotivados, tienen una percepción de baja competencia y no contingencia de los resultados académicos. A su vez los estudiantes con un tipo de regulación externa tiene una percepción del locus de causalidad externo. En el otro extremo los estudiantes intrínsecamente motivados tienen una mayor percepción de competencia y un locus de causalidad más interno. Estas observaciones, parecen sugerir, que el modelo SDT aporta una imagen en líneas generales consistente con el modelo atribucional del proceso de respuesta a los SET discutido en la sección anterior. Esto es, si se analizara el patrón atribucional en los estudiantes menos severos, para el modelo SDT, probablemente sería más próximo al patrón atribucional I (véase 4.2). Por el contrario, los estudiantes más severos de acuerdo al modelo SDT, tendrían un

patrón atribucional más próximo al II.

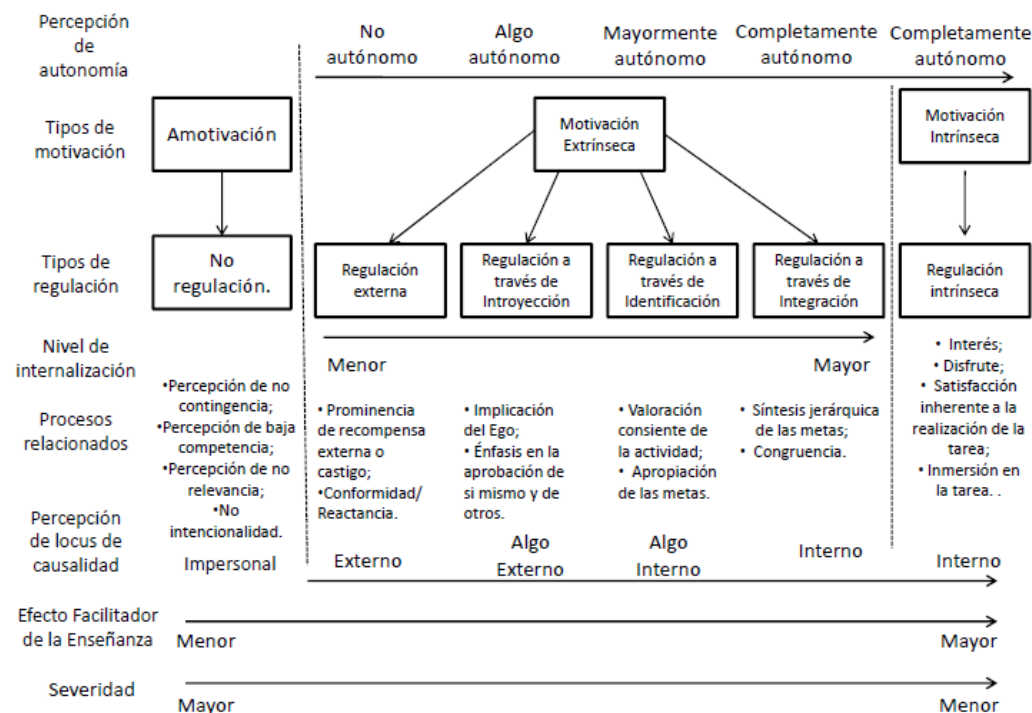


Figura 4.1: Modelo SDT del proceso de respuesta a los SET. Relación tipos de motivación y regulación con el efecto facilitador de la enseñanza y la severidad de los estudiantes.

Las características básicas del segundo componente del modelo se muestran en la Figura 4.2. Este componente explica la relación entre la satisfacción de necesidades básicas y las dimensiones del ambiente social, por un lado, y el efecto facilitador de la enseñanza y severidad de los estudiantes por otro.

De acuerdo al segundo componente del modelo, una enseñanza que ofrece un grado elevado de soporte a la autonomía, estructura (soporte a la competencia) y compromiso (soporte al relacionamiento), para un estudiante, tendría un efecto facilitador elevado del aprendizaje de este estudiante, que a su vez, sería poco severo al evaluar la enseñanza. Por el contrario, una enseñanza que ofrezca un bajo soporte a las necesidades psicológicas básicas, y en consecuencia, una baja satisfacción de estas necesidades, produciría un efecto facilitador bajo, y el estudiante tendría una severidad elevada al evaluar la enseñanza.

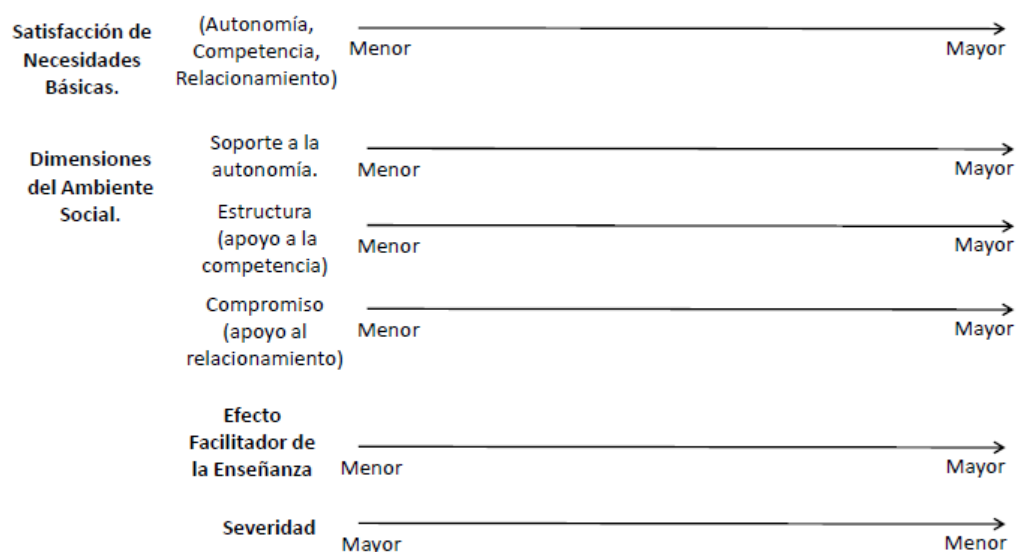


Figura 4.2: Modelo SDT del proceso de respuesta a los SET. Relación entre satisfacción de necesidades básicas, efecto facilitador de la enseñanza y severidad de los estudiantes.

Por último, en la Figura 4.3 se describe el tercer componente del modelo. Este componente explica la relación entre el tipo de contenido de metas y la satisfacción de necesidades psicológicas básicas, por un lado, y el efecto facilitador de la enseñanza y severidad de los estudiantes por otro. De acuerdo a este componente del modelo, los estudiantes que adoptan metas intrínsecas serían menos severos que estudiantes con metas extrínsecas. Como en los otros dos componentes anteriores, el efecto facilitador de la enseñanza tendría una relación inversa con la severidad. Las diferencias en el efecto facilitador de la enseñanza en el proceso de aprendizaje de los estudiantes, entre estudiantes que adoptan diferentes contenidos de metas, estaría vinculado a diferencias a nivel comportamental y cognitivo (Vansteenkiste y cols., 2006). La relación propuesta entre los contenidos de metas y la severidad de los estudiantes es consistente con evidencias previas (Svanum y Aigner, 2011).

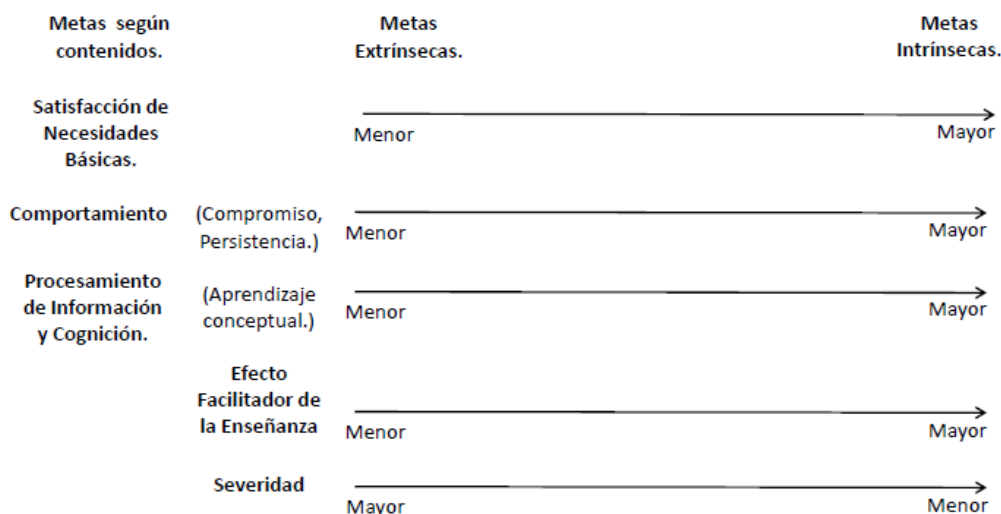


Figura 4.3: Modelo SDT del proceso de respuesta a los SET. Relación del contenido de metas con el efecto facilitador de la enseñanza y la severidad de los estudiantes.

4.4. Modelo TAS del proceso de respuesta a los SET.

En esta sección se describe un modelo del proceso de respuesta a los SET que sigue la perspectiva de la teoría del Aprendizaje Social de Bandura (TAS). En su conjunto permite establecer un vínculo entre los principales constructos de la TAS, por un lado, y el efecto facilitador de la enseñanza y la severidad de los estudiantes, por otro. A este modelo lo denominaremos modelo TAS del proceso de respuesta a los SET.

La Figura 4.4 sintetiza las principales características del modelo. El constructo fundamental en este modelo es la autoeficacia, que es la variable clave en la TAS. Otra variable importante es la expectativa de resultados, si bien, esta presenta un papel secundario. Como se aprecia en la Figura 4.4, el presente modelo afirma que estudiantes con una elevada autoeficacia serían menos severos al evaluar la enseñanza que estudiantes con una baja autoeficacia.

A su vez, el efecto facilitador de la enseñanza en un estudiante con una autoeficacia alta sería mayor. Este mayor efecto facilitador estaría relacionado con diferencias a nivel comportamental y cognitivo, vinculadas con diferencias en la autoeficacia (Bandura, 1989, 1991, 1993; Bandura y cols., 1996; Berry, 1999; Schunk, 1990).

Estudios previos aportan evidencias consistentes con el modelo propuesto. De acuerdo a [Wigfield y cols. \(1997\)](#) estudiantes con una mayor autoeficacia tendrían una percepción más positiva (menor severidad) al evaluar aspectos de la enseñanza tales como su valor, utilidad e importancia. Adicionalmente, los resultados aportados por [Wigfield y cols. \(1997\)](#) sugieren que los estudiantes con una elevada autoeficacia tenderían a percibir la enseñanza impartida por un docente en un curso como más interesante.

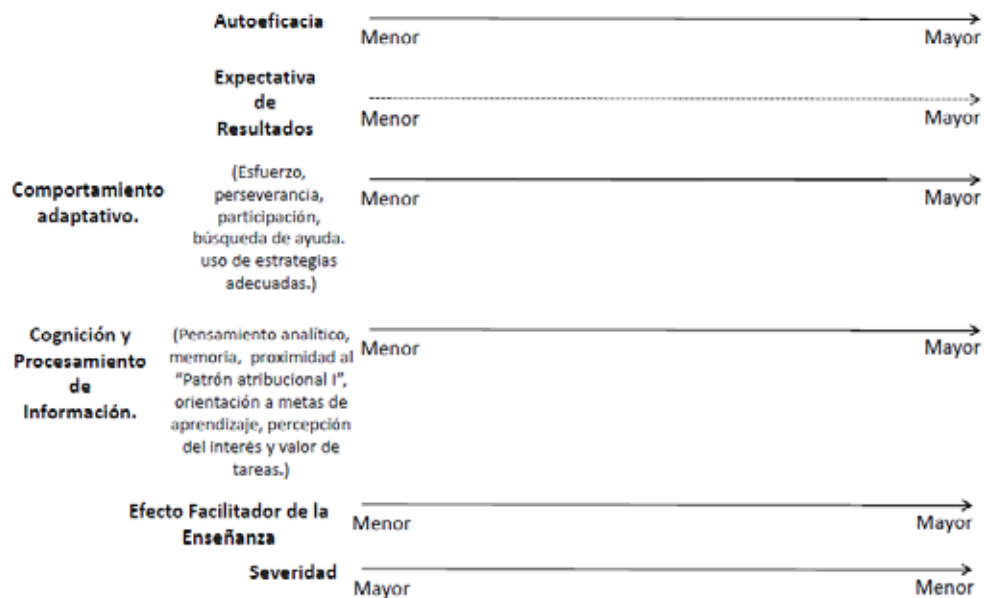


Figura 4.4: Modelo TAS del proceso de respuesta a los SET.

Por otra parte, es un aspecto central de la TAS que variables como la autoeficacia son variables de estado, cuyo valor resulta de interrelaciones entre factores personales y conductuales (de los estudiantes), por un lado, y factores ambientales (relacionados con la enseñanza), por otro. Dentro del marco de la TAS existe un gran cuerpo de evidencia que muestra que el grado en que la enseñanza promueve la autoeficacia está directamente relacionado con el efecto facilitador de la enseñanza sobre el aprendizaje y rendimiento académico ([Bandura y cols., 1996](#); [Zimmerman, 1995, 2000](#)).

A nivel global, este modelo parece también ser consistente con los modelos previamente discutidos. Así, una elevada autoeficacia está asociada a un orientación mayor a metas intrínsecas, que a metas extrínsecas ([Schunk, 1990](#)). Adicionalmente, una elevada autoeficacia estaría asociada a atribuciones causales más próximas al patrón atribucional I (véase 4.2). En consecuencia, estas

observaciones sugieren que no habrían inconsistencias subyacentes entre los modelos respecto a las predicciones realizadas sobre la severidad de los estudiantes.

4.5. Modelo de Auto-teorías del proceso de respuesta a los SET.

El modelo descrito en esta sección se basa en el modelo de Dweck y Leggett (véase sección 2.4.5). Este modelo permite establecer un vínculo entre los principales constructos considerados en el modelo de Dweck y Leggett, tales como el tipo de configuración mental adoptada (crecimiento o fija) y el tipo de meta (aprendizaje o desempeño), por un lado, y el efecto facilitador de la enseñanza y la severidad de los estudiantes, por otro. A este modelo se lo denominará modelo de auto-teorías del proceso de respuesta a los SET, haciendo referencia a un constructo fundamental en el modelo de Dweck y Leggett.

La Tabla 4.2 sintetiza las principales características del modelo. Como se aprecia en la Tabla 4.2, el modelo de auto-teorías del proceso de respuesta a los SET relaciona una menor severidad de los estudiantes y un mayor efecto facilitador de la enseñanza, con la adopción de una configuración mental de crecimiento. A su vez, el tipo de auto-teoría estaría relacionada con la selección de metas, de forma tal, que los sujetos con configuración mental de crecimiento tendrían una preferencia por metas de aprendizaje, respecto de metas de desempeño.

Adicionalmente, el efecto facilitador de la enseñanza depende también de la habilidad auto-percibida. La habilidad auto-percibida se mide con instrumentos que tienen contenidos afines y algunas características similares a los instrumentos utilizados para medir autoeficacia. A su vez, a nivel conceptual la habilidad auto-percibida tiene un fuerte solapamiento con el concepto de auto-eficacia. Por tanto, razonablemente, ambos constructos pueden ser identificados y discutidos como un mismo constructo.

La auto-eficacia tiene una relación de dependencia recíproca con las configuraciones mentales y los tipos de metas (Bandura, 1991; Schunk, 1990). En particular, la auto-eficacia tiende a aumentar en sujetos con configuraciones mentales de crecimiento y con metas de aprendizaje. Por tanto, los sujetos con una configuración mental fija y metas de desempeño tienden a mantener

o disminuir su autoeficacia. Por el contrario, los sujetos con una configuración mental de crecimiento y metas de aprendizaje tienden a aumentar su autoeficacia.

El tipo de configuración mental (crecimiento o fija) tiene una estrecha relación con las atribuciones causales de los estudiantes (Bandura, 1990, 1993; Wood y Bandura, 1989). Una mentalidad de crecimiento está vinculada a una concepción de la habilidad como inestable (que puede cambiar con el tiempo). Sujetos con una mentalidad de crecimiento consideran el esfuerzo y uso de estrategias causas importantes. Estos sujetos tienden a atribuir sus fracasos a falta de esfuerzo y estrategias inadecuadas. Adicionalmente, sus éxitos son atribuidos a un esfuerzo y estrategias, a su habilidad concebida como inestable, y tal vez, controlable, al menos, indirectamente a través del esfuerzo. Estos sujetos con una configuración mental de crecimiento identifican el esfuerzo como un medio para mejorar su habilidad.

Por tanto, el modelo de auto-teorías del proceso de respuesta a los SET considera simultáneamente varios de los constructos considerados por los otros modelos discutidos previamente: atribuciones causales, contenidos de metas y auto-eficacia. El patrón predicho de relaciones entre estos constructos, por un lado, y la severidad y el efecto facilitador de la enseñanza por otro, es consistente con los otros modelos. No obstante, el resumen de este modelo presentado en la Tabla 4.2 debe ser interpretado cuidadosamente considerando que subyacente a este modelo existe una dependencia recíproca multidimensional entre los constructos considerados.

Por último, el modelo de auto-teorías del proceso de respuesta a los SET relaciona los constructos fundamentales (atribuciones causales, contenidos de metas y auto-eficacia) con el procesamiento de información, respuestas emocionales y comportamiento. Por tanto, el modelo ofrece también un marco referencial para intentar entender las relaciones entre estos constructos con el efecto facilitador producido por la enseñanza y la severidad de los estudiantes, a niveles cognitivo, emocionales y comportamentales.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Tabla 4.2: Modelo de Auto-teorías del proceso de respuesta a los SET

| Auto-teoría | Orientación a Metas | Habilidad Auto-percibida | Procesamiento de Información | Respuestas Emocionales | Comportamiento | Efecto Facilitador de la Enseñanza | Severidad |
|---------------------------------------------------|--------------------------|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|-----------|
| Configuración Mental Fija (Entidad) | Metas de desempeño. | Baja | Relación inversa entre esfuerzo y habilidad. Las dificultades representan obstáculos. Enfrentar dificultades puede bajar la percepción de habilidad. Éxitos o fracasos se consideran no contingentes. | Ansiedad, afecto deprimido, aburrimiento, desobediencia, reveldía, evitación como respuesta a fracasos o necesidad de ejercer esfuerzo. Auto-valor y auto-estima dependen de contingencias afectivas y evaluativas. | No persistencia frente a dificultades, evitación de desafíos, baja o no participación, bajo o no esfuerzo. | Bajo | Elevada |
| Configuración Mental Fija (Entidad) | Metas de desempeño. | Alta | Relación inversa entre esfuerzo y habilidad. Las dificultades representan obstáculos. Enfrentar dificultades puede bajar la percepción de habilidad. | Auto-valor y auto-estima dependen de contingencias afectivas y evaluativas. | Elección de tareas que maximizan juicios positivos y reconocimientos de la habilidad; mientras que, minimizan juicios negativos, ansiedad y vergüenza. Persistencia dependiente de una alta percepción de habilidad. | Medio | Media |
| Configuración Mental de Crecimiento (Incremental) | Metas de aprendizaje,je. | Baja | El esfuerzo y estrategias permiten mejorar la habilidad. Información sobre dificultades, resultados, esfuerzo ejercido y estrategias empleadas son utilizadas para auto-evaluación y mejora. | Satisfacción, interés y disfrute por aprender. El esfuerzo ejercido puede verse acompañado de orgullo y disfrute. | Elección de tareas que puedan maximizar el desarrollo de la habilidad, interés y disfrute intrínseco al aprendizaje y dominio de una tarea. Esfuerzo y Persistencia. | Elevado | Baja |
| Configuración Mental de Crecimiento (Incremental) | Metas de aprendizaje,je. | Alta | El esfuerzo y estrategias permiten mejorar la habilidad. Información sobre dificultades, resultados, esfuerzo ejercido y | Satisfacción, interés y disfrute por aprender. El esfuerzo ejercido puede verse acompañado de orgullo y disfrute. | Elección de tareas que puedan maximizar el desarrollo de la habilidad, interés y disfrute intrínseco al aprendizaje y dominio de una tarea. | Más Elevado. | Más Baja. |

4.6. Discusión.

Existen diversas evidencias que sugieren que varios constructos importantes considerados en teorías socio-cognitivas de la motivación permiten explicar diferencias en la severidad de los estudiantes al evaluar la enseñanza. Estas evidencias son consistentes con las relaciones entre estos constructos y la severidad de los estudiantes, predichas por los modelos propuestos en este capítulo. [Svanum y Aigner \(2011\)](#) encontraron que estudiantes con metas intrínsecas evalúan más positivamente la enseñanza que aquellos con metas extrínsecas, en una evaluación al final del semestre de un curso universitario.

Por su parte, [R. Ames y Lau \(1979\)](#) encontraron que estudiantes con locus de causalidad interno evalúan más positivamente que aquellos con locus de causalidad externo. Un resultado similar fue encontrado por [Grimes y cols. \(2004\)](#) con el constructo locus de control. En este estudio se encontró que estudiantes con locus de control más interno evalúan más positivamente la enseñanza que estudiantes con locus de control externo.

Adicionalmente, [Wigfield y cols. \(1997\)](#) en un estudio longitudinal con niños de educación elemental, encontraron resultados consistentes con una relación positiva entre la autoeficacia y percepción del valor, utilidad, importancia e interés. Es decir, encontraron que una mayor autoeficacia se asoció a una percepción más positiva de la enseñanza (menor severidad).

Los estudios mencionados por tanto aportan evidencias de que los contenidos de metas, los patrones atribucionales y la autoeficacia permiten explicar diferencias en la severidad de los estudiantes. Aún más, la dirección de los efectos es consistente con los modelos propuestos.

Por otra parte, cuando se observan las relaciones entre varios de los constructos considerados se encuentra un patrón consistente de los cuatro modelos propuestos con las evidencias, y entre los modelos considerados, en relación a las predicciones sobre la severidad esperada de los estudiantes al evaluar la enseñanza. Por ejemplo, las relaciones entre configuraciones mentales y patrones atribucionales con la adopción de diferentes contenidos de metas ([Bandura, 1990, 1993; Wood y Bandura, 1989](#)) permiten deducir la consistencia entre el modelo atribucional, el tercer componente del modelo SDT y el modelo de auto-teorías, entre sí, y con las evidencias antes mencionadas. En forma similar, la consistencia entre modelos y las evidencias previas puede deducirse considerando la relación entre la autoeficacia y contenidos de metas ([Schunk,](#)

1990).

En particular, si se consideran los tres componentes del modelo SDT, se puede apreciar un patrón subyacente común. Cuanto mayor es la satisfacción de las necesidades psicológicas básicas, mayor es el efecto facilitador de la enseñanza, y menor, la severidad de los estudiantes al evaluar la enseñanza. Este patrón es altamente consistente con el papel que tienen las necesidades psicológicas básicas como principio unificador en la SDT.

De los anteriores argumentos, una primer conclusión, es que los modelos propuestos parecen ofrecen una imagen general coherente del proceso de respuesta a los SET, desde diferentes perspectivas teóricas, que a su vez, esta en línea con resultados encontrados en varios estudios previos. Así, los modelos propuestos podrían ofrecer una buena base para el desarrollo de una teoría socio-cognitiva del proceso de respuesta a los SET, que vincule constructos fundamentales de las cuatro perspectivas teóricas consideradas, con el efecto facilitador de la enseñanza y la severidad de los estudiantes.

Globalmente, los modelos propuestos aportan una visión general de como varios constructos importantes median la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza y como se producen diferencias entre los estudiantes en como la enseñanza facilita su aprendizaje.

Adicionalmente, los modelos propuestos en su conjunto orientan en la comprensión de como el efecto facilitador de la enseñanza y las diferencias en la severidad de los estudiantes, se asocia a diferencias a nivel cognitivo, afectivo/emocional y comportamental, relacionadas con los constructos fundamentales considerados en estos modelos.

Además, los modelos propuestos ayudan a comprender mejor cuando las diferencias observadas entre los estudiantes al evaluar la enseñanza son estrictamente sesgos, o alternativamente, reflejan diferencias individuales en el efecto facilitador de la enseñanza en relación al proceso de aprendizaje de los estudiantes.

Por tanto, una segunda conclusión que parece razonable formular a partir de estos argumentos, es que los modelos propuestos en este capítulo aportan un marco teórico útil para el estudio de la validez de constructo de los SET basado en el estudio teórico y empírico del proceso de respuesta de los estudiantes al evaluar la enseñanza, desde una perspectiva cognitiva y socio-cognitiva.

Por otra parte, en la actualidad predomina en las prácticas de evaluación de la calidad de la enseñanza una orientación hacia la enseñanza, y no, una

orientación centrada en el estudiante y el aprendizaje, más consistente con tendencias actuales en las prácticas de enseñanza (Theall y cols., 2001; Weimer, 2002).

Los modelos propuestos aquí podrían aplicarse para implementar una evaluación de la calidad de la enseñanza centrada en estudiante y su aprendizaje. Esto podría implementarse midiendo el efecto de la enseñanza en el proceso de aprendizaje de los estudiantes, mediante la medición del efecto en los constructos considerados en estos modelos. Así, una tercera conclusión, es que estos modelos darían un marco conceptual adecuado para prácticas de evaluación formativa de la calidad en la enseñanza y el aprendizaje de los estudiantes.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Capítulo 5

Diagnóstico de modelos de múltiples facetas de Rasch en estudios observacionales.

5.1. Presentación.

El modelo de múltiples facetas de Rasch (MFRM) tiene aplicaciones en diversas áreas (Bond y Fox, 2015; Engelhard Jr, 2013; Linacre, 2015). MFRM permite modelar la respuesta a ítems de una escala como un fenómeno probabilístico influenciado por elementos de varias facetas. Cada faceta corresponde a factores que afectan la probabilidad de la respuesta descrita mediante una variable ordinal. Un aspecto clave en la aplicación del MFRM es la conexión entre los elementos de las facetas. Este aspecto del diseño afecta la exactitud y precisión de las medidas y la aplicación de los procedimientos de diagnóstico del modelo.

En estudios experimentales la conexión es definida a priori. No obstante, en estudios observacionales la conexión emergente puede ser compleja y deficitaria. Por tanto, se requieren métodos especializados de diagnóstico con los que actualmente no se cuenta.

Los objetivos de este trabajo son: (a) caracterizar la conexión emergente en un estudio observacional; (b) comparar un método de diagnóstico corrientemente usado (Linacre, 2015), con un método alternativo.

El modelo MFRM utilizado aquí incluyó tres facetas de medición (docentes, estudiantes e ítems). Para caracterizar la conexión emergente se aplicó un

modelo de gráfica no dirigida (Trudeau, 2013) y se determinó el tipo de red de evaluación (Engelhard Jr, 1996). En la comparación de los métodos de diagnóstico se consideró: la estabilización de las mediciones; el ajuste global; y las medidas de invarianza.

La conexión en la faceta docentes resultó mayor que la observada en la faceta estudiantes. La red de evaluación encontrada resultó ser una red de evaluación conectada incompleta e incluyó conjuntos de elementos que son claves para la conexión de la red. Los criterios de comparación considerados sugieren que el método propuesto funciona mejor.

Respecto al problema central abordado en esta tesis, el método diagnóstico propuesto resulta mucho más adecuado debido a que permite considerar el proceso de respuesta de un alto porcentaje de los evaluadores (estudiantes) en la definición de la escala de medida (90 % en comparación con un 40 %).

5.2. El Modelo de medición de múltiples facetas de Rasch

El modelo de medición de múltiples facetas de Rasch (modelo MFRM) utilizado para analizar los datos de respuestas a escalas de evaluación de la efectividad en la enseñanza puede ser especificado como sigue:

$$\ln \left[\frac{P_{nij k}}{P_{nij k-1}} \right] = \theta_n - \beta_i - \alpha_j - \tau_{ik} \quad (5.1)$$

donde

$P_{nij k}$ = Probabilidad de que el docente n reciba un puntaje k en el ítem i del estudiante j ,

$P_{nij k-1}$ = Probabilidad de que el docente n reciba un puntaje $k-1$ en el ítem i del estudiante j ,

θ_n = Habilidad del docente n en la dimensión considerada,

β_i = Dificultad del ítem i ,

α_j = Severidad del estudiante j ,

τ_{ik} = Dificultad de recibir un puntaje k relativo a recibir un puntaje $k-1$ para el ítem i .

Este modelo es un modelo MFRM de crédito parcial (Linacre, 2014). Puede considerarse una extensión del modelo de crédito parcial (Masters, 1982) en el que se considera el efecto del evaluador en las respuestas (Linacre, 2014). El modelo de crédito parcial podría especificarse como un caso particular del modelo MFRM, donde únicamente se consideran la faceta docente y la faceta items, con umbrales entre las categorías que pueden diferir para cada ítem. En el modelo MFRM de crédito parcial considerado se incluye además la faceta estudiante. La faceta estudiante permite medir la severidad de cada estudiante al evaluar al docente.

Masters y Wright (1997) discuten la relación del modelo de crédito parcial con otros modelos unidimensionales para respuestas con dos o más respuestas ordenadas, mientras que Linacre (2014) explica las diferencias entre el modelo MFRM de crédito parcial y el MFRM de escala de puntajes, que es a su vez una extensión del modelo de escala de puntajes (Andrich, 1978).

La ecuación (7.1) especifica un modelo de medición con tres facetas que definen los efectos primarios en el proceso de respuesta. Además de las tres facetas consideradas en este modelo de medición, se definieron dos facetas dummy, que contienen elementos con medida cero, y por tanto, no afectan la medida de los elementos de las otras facetas.

Estas facetas dummy se definieron para poder particionar el ajuste, obtener medidas de ajuste global y para poder describir y comparar las medidas de los docentes de diferentes cursos. La aproximación empleada es descrita por Linacre (2014).

La primer faceta dummy se denominó curso. Esta faceta representa las actividades de enseñanza llevadas cabo por los docentes. La faceta curso incluyó 10 elementos. La segunda faceta dummy, se definió para obtener estadísticos de ajuste globales (Infit y Outfit para el modelo) y contiene un único elemento presente en todas las respuestas.

5.3. Muestra de datos de calibración.

La muestra de datos de calibración considerada aquí consistió en las respuestas a la sub-escala VALR (E1) del cuestionario de evaluación estudiantil de la calidad en la enseñanza de 3243 estudiantes, que evaluaron 98 docentes, en 10 actividades de enseñanza, durante dos semestres consecutivos, en la carrera de médico, durante el año 2015.

5.4. Diagnóstico de ajuste entre datos y el modelo.

Una aproximación en el proceso de diagnóstico es eliminar los elementos con falta de ajuste, estimar nuevamente las medidas de los elementos y luego comparar las estimaciones obtenidas con las estimaciones previas. Si se produce un cambio importante en las estimaciones, entonces se eliminan los elementos que ahora aparecen con falta de ajuste. Este proceso se repite mientras se encuentre un cambio perceptible en las estimaciones.

En este proceso se considera además si las medidas mejoran considerando las medidas de ajuste que indican el grado de ajuste entre cada elemento y el modelo (Linacre, 2010). Adicionalmente, Linacre (2014) recomienda que en primer lugar se eliminen respuestas con residuos extremos y elementos con correlaciones punto-biseriales negativas.

Como resultado de ese proceso de diagnóstico se espera que los elementos que permanecen mejoren su ajuste y que las estimaciones de las medidas se estabilicen. No obstante, con los datos analizados, esto no ocurrió, incluso eliminando una parte substancial de los elementos de la faceta estudiantes. Esto, a pesar que el ajuste global inicial resultó próximo a un nivel aceptable y que el ajuste en las facetas docente e ítems resultó aceptable, con todos su elementos a un nivel aceptable de ajuste.

Debido al problema señalado, se decidió adoptar un aproximación diferente, en la cual se comienza por eliminar solamente elementos con un muy bajo ajuste presentes en respuestas con residuos altamente extremos. En este aspecto, esta aproximación más refinada y gradual es similar a la adoptada por Farrokhi y Esfandiari (2011).

5.5. Método de diagnóstico alternativo propuesto.

Para implementar el proceso de diagnóstico se definió un algoritmo que considera las cantidades: valores absolutos de los residuos estandarizados (AbsStRes), InfitMS, OutfitMS, InfitZ y OutfitZ. Además, se considero el porcentaje de cambio en la medida estimada de cada elemento. El porcentaje de cambio en la medida estimada de cada elemento de una faceta respecto a su estimación

en el paso anterior de calibración fue definido como:

$$\% \Delta M_n = 100 \cdot \frac{\|M_n - M_{n-1}\| + \epsilon}{\|M_{n-1}\| + \epsilon} \quad (5.2)$$

donde $0 < \epsilon < 1$. En particular, se realizaron los cálculos con $\epsilon = 0,01$. $\% \Delta M_n$ es la razón porcentuada entre la diferencia entre dos pasos consecutivos $n-1$ y n en valor absoluto, y el valor de la medida estimada en el paso $n-1$. La constante ϵ que se suma en el numerador y denominador permite que $\% \Delta M_n$ esté definida incluso cuando M_{n-1} vale cero. Para valores de M_{n-1} cercanos a cero, $\% \Delta M_n$ es mayor que para valores más alejados de cero. Por esa razón, para analizar la estabilidad de las estimaciones también se definió la cantidad:

$$\delta M_n = \|M_n - M_{n-1}\| \quad (5.3)$$

δM_n expresa la variación absoluta en la medida entre dos pasos consecutivos $n-1$ y n en unidades logits.

Adicionalmente, el algoritmo definido considera una serie de etapas con criterios para definir respuestas extremas y criterios de eliminación de elementos con falta de ajuste. En la Tabla 5.1 se definen los criterios adoptados para cada etapa en el proceso de diagnóstico de falta de ajuste.

Observar que los elementos eliminados en cada etapa deben verificar alguno de los criterios de eliminación de elemento con falta de ajuste. Pero también, debe pertenecer a una respuesta extrema, definida por el criterio para la etapa correspondiente. Esto implica que no todos los elementos con falta de ajuste son eliminados en cada paso del proceso de diagnóstico.

Por otra parte, cada etapa de diagnóstico comprende una secuencia de pasos hasta que no existen más respuestas y elementos que verifiquen las condiciones de la etapa. Cuando eso ocurre se pasa a la etapa siguiente. En cada etapa, el número de elementos eliminados disminuye hasta que ningún elemento verifica las condiciones definidas para dicha etapa.

También, se definió el criterio adoptado para parar el proceso de diagnóstico. Así, se detiene la eliminación de elementos con falta de ajuste cuando el tercer cuartil de los porcentajes de cambio en medidas estimadas es menor o igual a 2,5% para los elementos de todas las facetas consideradas en dos pasos consecutivos. En consecuencia, mientras no se verifique el criterio de parada, cuando no hay elementos que verifiquen las condiciones de respuesta extrema y de eliminación de elemento con falta de ajuste, se pasa a la siguiente etapa

de diagnóstico.

Tabla 5.1: Etapas en el proceso de diagnóstico de falta de ajuste.

| | Respuesta extrema | Elemento con falta de ajuste |
|---------|-------------------|----------------------------------------------------------------------|
| Etapa 1 | $AbsStRes > 3$ | $InfitMS > 2$ y $InfitZ > 2$ o $OutfitMS > 2$ y $OutfitZ > 2$ |
| Etapa 2 | $AbsStRes > 3$ | $InfitMS > 1,5$ y $InfitZ > 2$ o $OutfitMS > 1,5$ y $OutfitZ > 2$ |
| Etapa 3 | $AbsStRes > 2,5$ | $InfitMS > 1,5$ y $InfitZ > 2$ o $OutfitMS > 1,5$ y $OutfitZ > 2$ |
| Etapa 4 | $AbsStRes > 2$ | $InfitMS > 1,5$ y $InfitZ > 2$ o $OutfitMS > 1,5$ y $OutfitZ > 2$ |

5.6. Resultados.

5.6.1. Caracterización de la muestra de datos de calibración.

Dos características importantes de la muestra de calibración son el número de observaciones de cada elemento y el solapamiento o conexión entre estos elementos. Los ítem están completamente conectados con todos los estudiantes y docentes considerados en la presente muestra. Por el contrario, la conexión es menor en docentes y estudiantes. En esta sección se muestran los resultados que se obtuvieron de la caracterización de la muestra de datos de calibración.

Número de observaciones.

En la Figura 5.1 se muestra el número de observaciones de cada elemento de las tres facetas consideradas para las escalas sobre la enseñanza. El número de observaciones de docentes y estudiantes varía según el número de ítems de la escala considerada. En la Figura 5.1 se muestra el número de observaciones para escalas de seis ítems y de doce ítems. La longitud de seis ítems coincide con la longitud promedio de las escalas sobre la enseñanza. La longitud de doce ítems corresponde aproximadamente a la longitud de las escalas de medición de factores de segundo orden de la enseñanza (Marsh y Hocevar, 1985).

En la Figura 5.1 se puede apreciar que el número de observaciones de los elementos de la faceta docentes e ítems es siempre superior al límite sugerido

por [Linacre \(2014\)](#) como orientación para obtener medidas con una buena precisión. En el caso de los estudiantes existen elementos que tienen un número de observaciones más pequeño. Un porcentaje importante de los estudiantes tiene un número de observaciones menor a treinta, por lo que se espera una menor precisión en la estimación de sus medidas.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



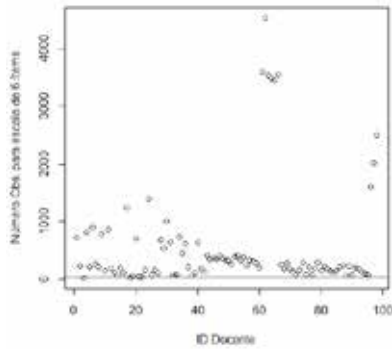
comisión sectorial
de enseñanza



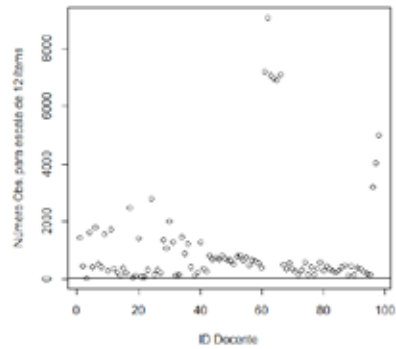
Facultad
de Humanidades
y Ciencias de la Educación



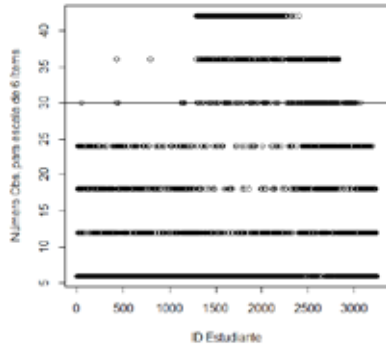
Consejo de
Formación en
Educación



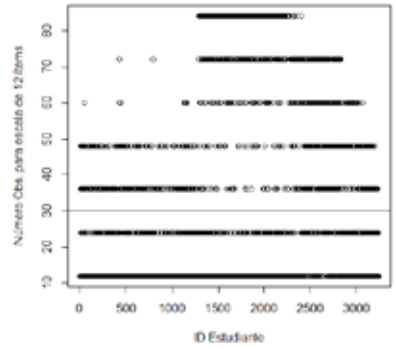
(a) Número Obs. de Docentes - Escala de 6 ítems



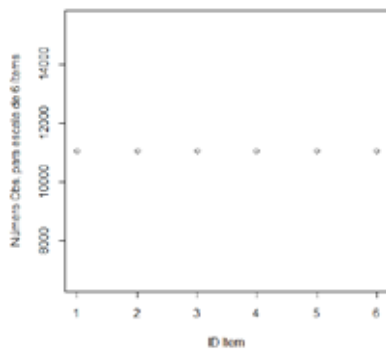
(b) Número Obs. de Docentes - Escala de 12 ítems



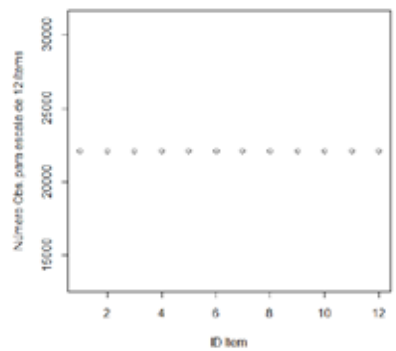
(c) Número Obs. de Estudiantes - Escala de 6 ítems



(d) Número Obs. de Estudiantes - Escala de 12 ítems



(e) Número Obs. de Ítems - Escala de 6 ítems



(f) Número Obs. de Ítems - Escala de 12 ítems

Figura 5.1: Número de observaciones de elementos de cada faceta

Conexión.

En las Figuras 5.3 y 5.4 se describe la conexión de los elementos de las facetas docentes y estudiantes. Para ello se aplican conceptos básicos de teoría de gráficas (Trudeau, 2013). Así, se representa la conexión entre los elementos de las facetas docentes y estudiantes como una gráfica no dirigida (véase Figura 5.2), donde cada elemento es un vértice de la gráfica y cada borde representa que dos elementos están directamente conectados (son adyacentes). Esto ocurre en este caso cuando dos elementos se encuentran en una misma observación.

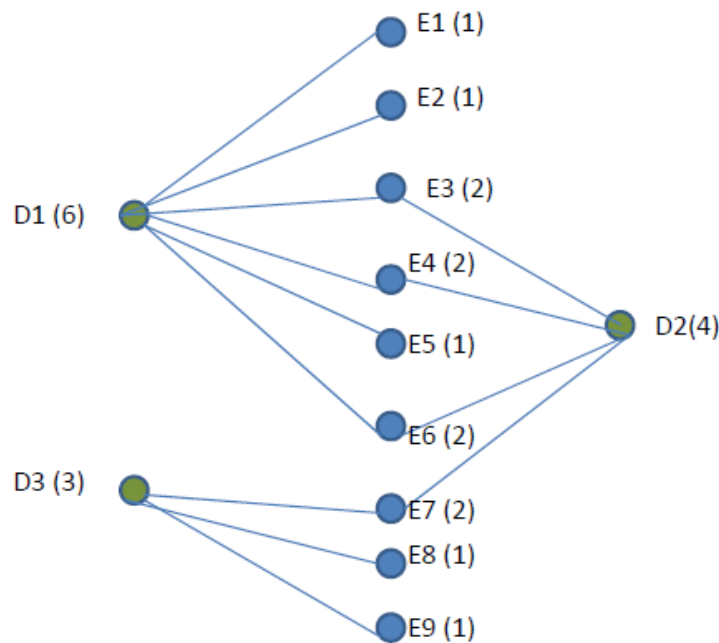


Figura 5.2: Gráfica de la faceta docentes y estudiantes. D1, D2 y D3 representan docentes evaluados, mientras que, E1 a E9 representan estudiantes. Los bordes de la gráfica unen los elementos (vértices) conectados entre sí. Entre paréntesis se indica el grado o número de elementos adyacentes o conexiones directas que tiene cada elemento.

El esquema de la Figura 5.2 permite apreciar que además de las conexiones directas entre los elementos de las facetas, también existen conexiones indirectas. Por ejemplo, el estudiante E1 está conectado indirectamente con el estudiante E2 por un camino de longitud dos, que incluye el borde {E1,D1} y el borde {D1,E2}. Mientras que el estudiante E1 está conectado con el estudiante E9 mediante un camino de longitud seis. En forma similar, se puede

apreciar que el docente D1 está conectado con el docente D2 por un camino de longitud dos, mientras que el docente D1 está conectado con el docente D3 mediante un camino de longitud cuatro.

En teoría de gráficas las conexiones directas entre los vértices de una gráfica se describen mediante la matriz de adyacencia (A). En esta matriz cada entrada corresponde a un par de vértices y toma valor uno cuando los vértices son adyacentes (están conectados por el borde que definen) o cero si no están directamente conectados. Por otro lado, las conexiones indirectas se pueden describir mediante la potencia n -ésima de la matriz A . La potencia n -ésima de la matriz A corresponde en cada entrada de la matriz al número de caminos de longitud n que conectan un par de vértices.

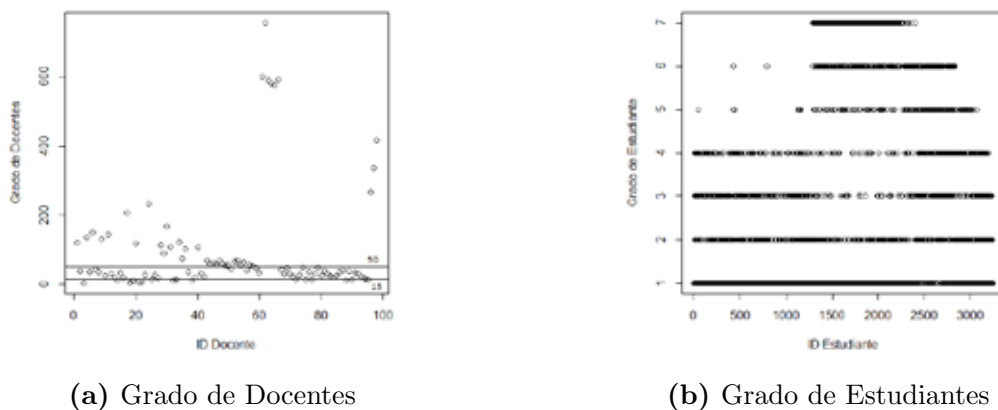


Figura 5.3: Conexión entre los elementos de las facetas Docentes y Estudiantes. El grado corresponde al número de elementos al que está conextado cada elemento coniderado.

En la Figura 5.3 se describe la conexión de los elementos de las facetas considerando el grado de los vértices. El grado de un vértice corresponde al número de vértices adyacentes a él o directamente conectados. En la Figura 5.3 se aprecia que el grado de los docentes en general es mayor al grado de los estudiantes. Esto significa que en la faceta estudiantes los elementos resultaron estar menos conectados que en la faceta docentes. Esto implica que se espera una menor exactitud en la estimación de las medidas de los estudiantes que en las medidas de los docentes.

Adicionalmente, una descripción más detallada de la conexión se puede hacer considerando la matriz A y sus primeras potencias. En la Figura 5.4, para cada entrada de las primeras seis potencias de A se traza un punto cuando

la entrada es mayor que cero. Además, en la matriz A las primeras 3243 filas y columnas corresponden a elementos de la faceta estudiantes. Por otro lado, las filas y columnas 3244 a 3341 corresponden a elementos de la faceta docentes.

Mientras que el gráfico correspondiente a la matriz A representa los bordes que conectan vértices adyacentes, los gráficos de las matrices A^n muestran si existen caminos de longitud n que unen pares de vértices. Así, mientras que la imagen correspondiente a la matriz A muestra la conexión entre elementos de facetas distintas, esto es, elementos de la faceta estudiantes con elementos de la faceta docentes, las matrices A^n muestran también la conexión entre elementos de una misma faceta.

En la Figura 5.4b se aprecia que los elementos de la faceta docentes presentan una conexión elevada entre ellos, mientras que la conexión entre los elementos de la faceta estudiantes es menor. En la Figura 5.4b también se aprecia que existen cuatro bloques diagonales. Los primeros tres corresponden a estudiantes y el último a docentes. Dentro de los tres bloques correspondientes a los estudiantes existe una mayor conexión que entre bloques. En particular el segundo bloque de estudiantes está mucho menos conectado a los otros dos.

A pesar de ello, en las Figuras 5.4c a 5.4f se aprecia que existen caminos de mayor longitud que conectan todos los elementos considerados y no existen subconjuntos de elementos desconectados. Este resultado tiene un significado importante en relación a la estimación de las medidas de los elementos y la identificación no ambigua del modelo de medición.

Para interpretar este resultado, es útil considerar el concepto de red de evaluación. Engelhard Jr (1996) define esencialmente una red de evaluación como un sistema interconectado de elementos de diferentes facetas. Estas redes de evaluación son completas cuando todos los elementos están conectados, incompletas cuando no todos los elementos están conectados, y no conectadas cuando existe subconjuntos de elementos no conectados. Por tanto, en este caso, la red de evaluación obtenida es una red conectada incompleta. De esta forma, es posible estimar las medidas de los elementos directamente y sin necesidad de adoptar asunciones adicionales para conectar subconjuntos de elementos no conectados (Engelhard Jr, 1996).

No obstante, en este caso, la conexión en uno de los bloques diagonales de estudiantes en la Figura 5.4b con los otros bloques depende de un número relativamente bajo de estudiantes. Esto implica que en el proceso de calibración

y diagnóstico debe considerarse la posibilidad de una pérdida importante de conexión en caso de eliminar una parte de estos elementos. Esta pérdida, podría significar una menor exactitud en la estimación de las medidas de los elementos y cierta distorsión en las medidas diagnósticas.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

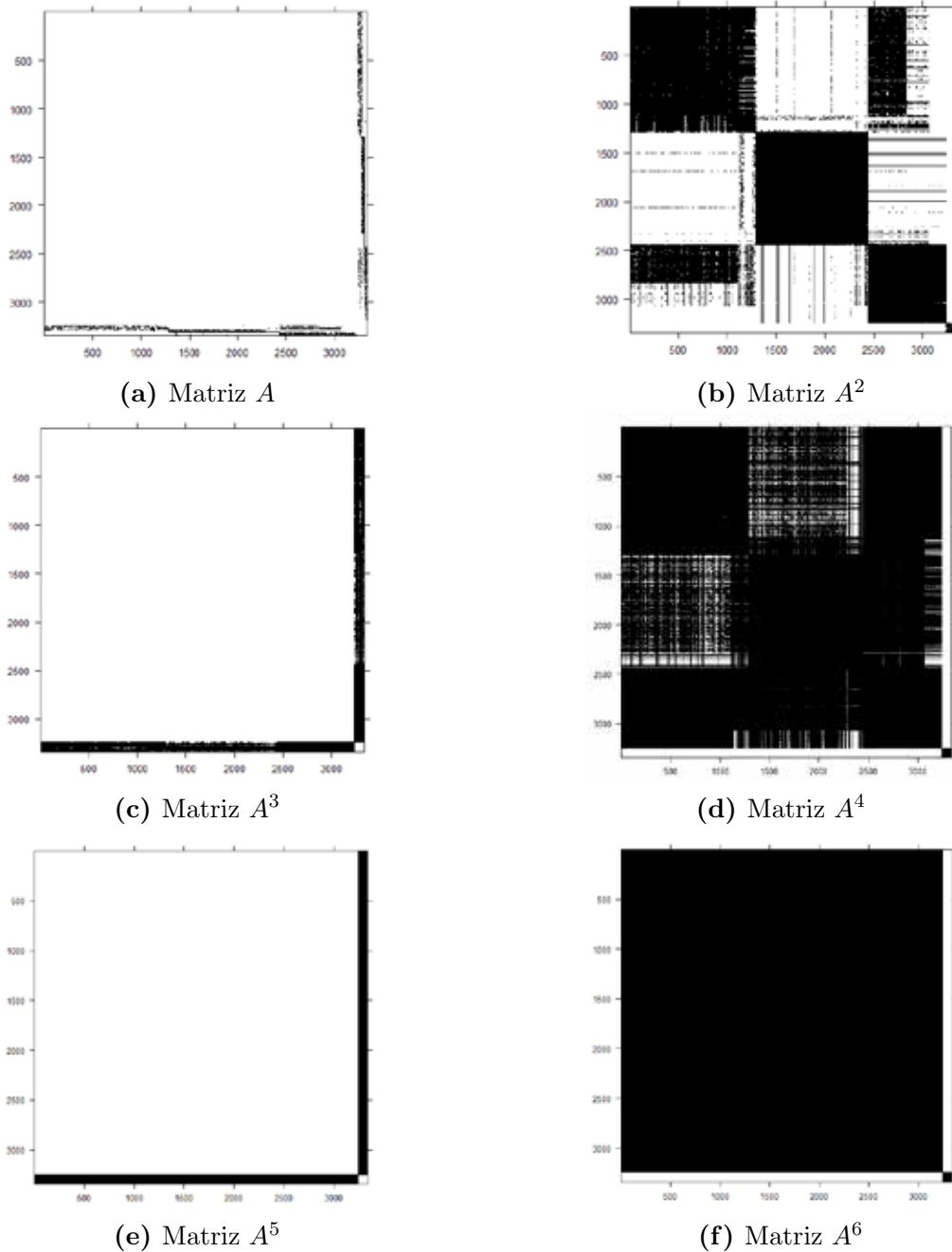


Figura 5.4: Conexión entre los elementos de las facetas Docentes y Estudiantes. Se muestran las entradas no nulas para la matriz de adyacencia A y sus primeras potencias. Cada punto corresponde a una entrada no nula. Las primeras 3243 filas y columnas corresponden a estudiantes y las últimas 98 filas y columnas corresponden a docentes.

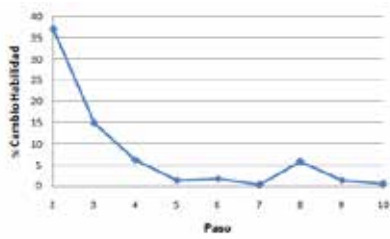
5.6.2. Evaluación empírica del método de diagnóstico.

Análisis de la estabilización de las medidas.

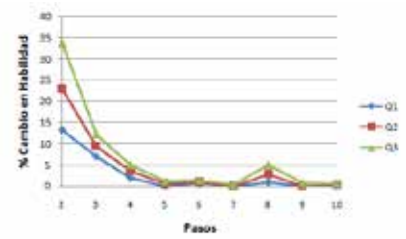
En este apartado se evalúa empíricamente el método de diagnóstico propuesto. Debido a que el problema de conexión es exactamente el mismo para los datos de todas las sub-escalas del cuestionario considerado, se eligió una en particular para hacer esta evaluación. Los resultados obtenidos sobre la evaluación empírica del método de diagnóstico corresponden a los datos de la sub-escala valor.

Si el algoritmo funciona adecuadamente, se esperaría encontrar que: (a) las medidas estimadas de los elementos de todas las facetas deberían estabilizarse a medida que se avanza en los pasos y etapas del algoritmo de diagnóstico; (b) el ajuste global entre los datos y el modelo debería mejorar a medida que se avanza en los pasos y etapas del algoritmo de diagnóstico; (c) el ajuste global final debería ser similar o mejor al que se obtiene de la aplicación estricta del algoritmo sugerido por Linacre, considerando el número de elementos con falta de ajuste eliminados; (d), la invarianza de las medidas debería mejorar a medida que se avanza en los pasos y etapas del algoritmo diagnóstico; y (e) la invarianza final de las medidas debería ser similar o mejor que la invarianza de las medidas obtenidas cuando se aplica el algoritmo que se apega estrictamente a las recomendaciones de Linacre.

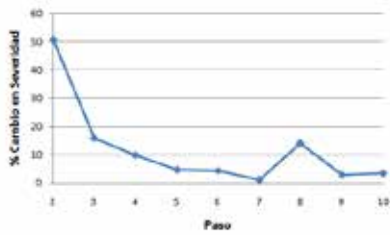
En la Figura 5.5 se muestra la media y cuartiles de $\% \Delta M_n$ para los elementos de las tres facetas consideradas. También, se muestra la media y cuartiles de δM_n para la faceta ítems. La Figura 5.5 muestra que efectivamente las medidas estimadas de los elementos de todas las facetas consideradas se estabilizan. El criterio de parada se verificó en el paso diez en la etapa 3. No obstante, a partir del paso cinco se alcanza un alto nivel de estabilidad en las estimaciones. Esto se aprecia también cuando se observa el comportamiento de las estimaciones para cada elemento de las facetas (Figura 5.6).



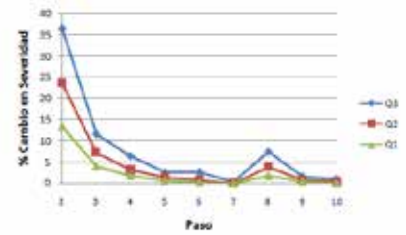
(a) Media del %Δ Habilidad



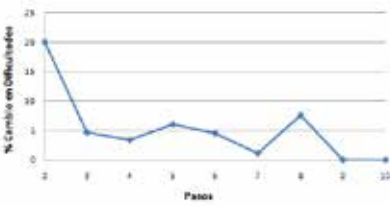
(b) Cuartiles del %Δ Habilidad



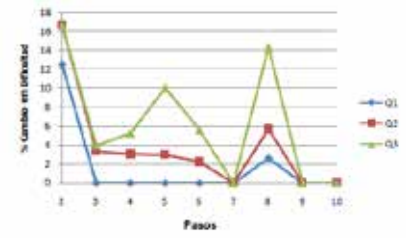
(c) Media del %Δ Severidad



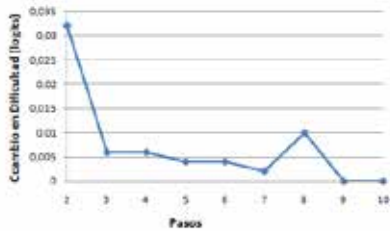
(d) Cuartiles del %Δ Severidad



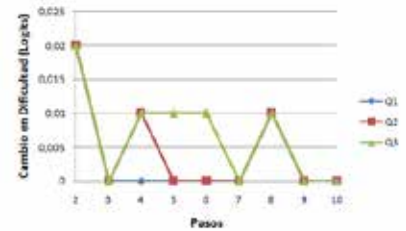
(e) Media del %Δ Dificultad



(f) Cuartiles del %Δ Dificultad

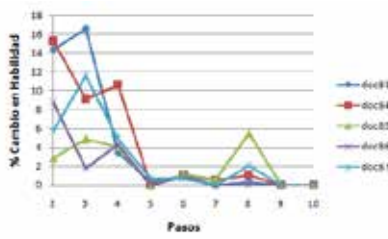


(g) Media de δ Dificultad

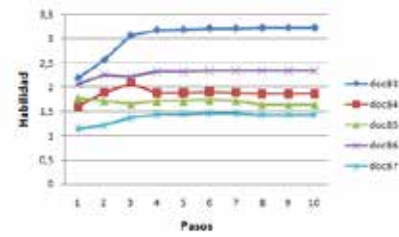


(h) Cuartiles del δ Dificultad

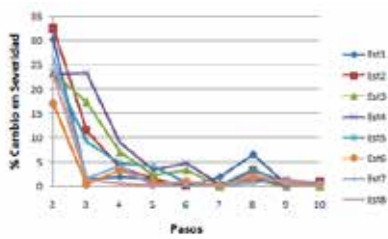
Figura 5.5: Estabilidad de las estimaciones en función de n para estimaciones de Habilidad, Severidad y Dificultad.



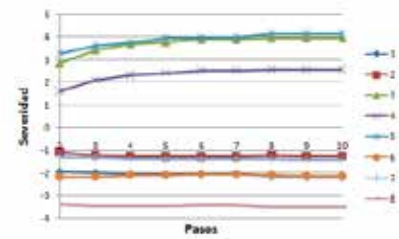
(a) Ejemplo: %Δ Habilidad



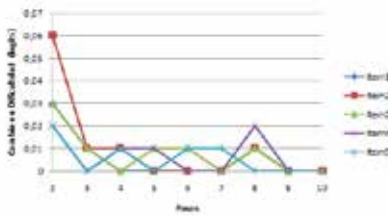
(b) Estimaciones de Habilidad en función de n



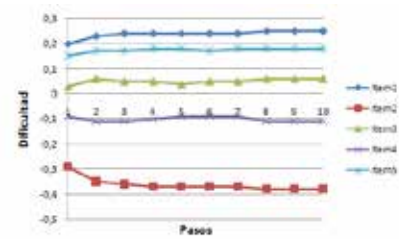
(c) Ejemplo: %Δ Severidad



(d) Estimaciones de la Severidad en función de n



(e) δ Dificultad



(f) Estimaciones de la Dificultad en función de n

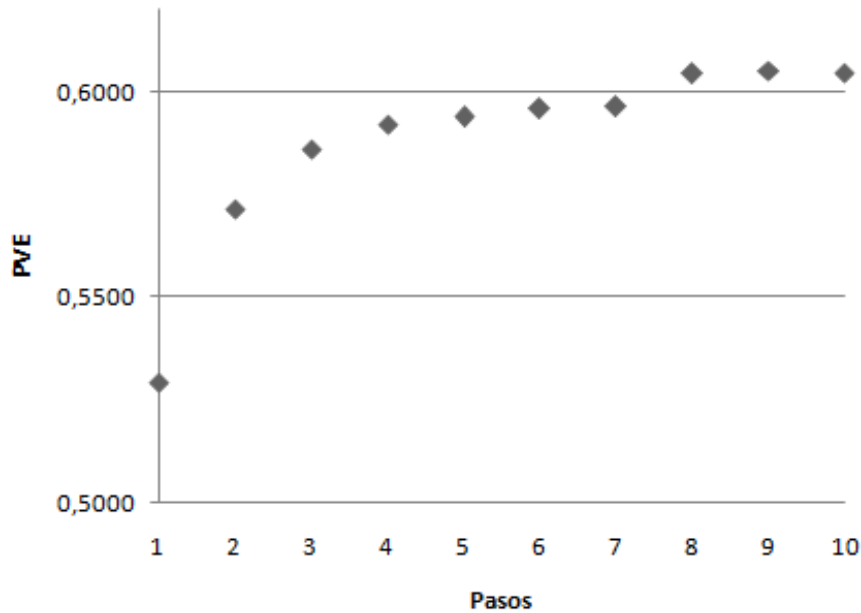
Figura 5.6: Ejemplos de estabilidad para elementos en cada faceta en función de n .

Ajuste global entre datos y el modelo.

Como se mencionó antes, si el algoritmo diagnóstico propuesto funciona adecuadamente, se espera que el ajuste entre los datos y el modelo mejore, al incrementarse el número de pasos. La Figura 5.7 muestra el valor de PVE en función del paso de diagnóstico n . Como se puede apreciar el valor de PVE aumenta en cada paso de diagnóstico hasta estabilizarse en un valor superior a 0,60. Esto indica una mejora en el ajuste absoluto entre los datos y el modelo. Este cambio en el ajuste global también se acompaña de una disminución en la falta de ajuste entre los elementos de cada faceta y el modelo. Esto es indicado por la disminución en el número de elementos que verifican la condición de

falta de ajuste en cada etapa.

Figura 5.7: PVE para el modelo en cada paso de eliminación de elementos con falta de ajuste



En la Figura 5.7 se aprecia también que la mejora en el ajuste global sugerida por el valor de *PVE* es mayor en los primeros pasos del algoritmo diagnóstico. A partir del paso 5 la mejora en el ajuste global resulta más lenta en los pasos siguientes hasta el paso 10.

En la Tabla 5.2 se muestra información detallada sobre el ajuste global para los pasos 1, 5 y 10 (método de diagnóstico alternativo) y para el método de diagnóstico de Linacre. A partir de la comparación de los valores mostrados en la Tabla 5.2 se aprecia una mejora importante entre el ajuste inicial y el paso 5. No obstante, la diferencia es menor entre el paso 5 y el paso 10. En forma similar, la diferencia es también menor cuando se compara entre el ajuste global en el paso 5 del método alternativo y el ajuste final obtenido siguiendo estrictamente el procedimiento propuesto por Linacre.

Para comparar las medidas de ajuste es importante considerar que el ajuste global del modelo tiene una relación proporcional con el número de elementos con falta de ajuste eliminados. Observar que eliminando tan solo un 10,2% de los estudiantes se observa en el paso 5 del algoritmo alternativo un $PVE = 0,594$; mientras que, cuando se sigue estrictamente el algoritmo sugerido por

Linacre se obtiene un PVE un 14% mayor que resulta de eliminar un 60,2% de los estudiantes originalmente considerados. Esto sugiere que el algoritmo propuesto permite obtener un nivel comparable de ajuste, pero, considerando aproximadamente un 90% de los estudiantes originalmente considerados, en lugar de un 40% que se considera en este caso siguiendo estrictamente el algoritmo sugerido por Linacre.

Tabla 5.2: Comparación del ajuste global

| | Ajuste Inicial | Paso 5 | Paso 10 | Ajuste Linacre |
|----------------------------|----------------|--------|---------|----------------|
| <i>Infit</i> | 0,990 | 0,990 | 0,990 | 0,990 |
| <i>Outfit</i> | 1,09 | 0,980 | 0,980 | 1,03 |
| <i>Media</i> (Z_{Res}) | -0,003 | 0,002 | 0,001 | 0,008 |
| <i>SD</i> (Z_{Res}) | 1,016 | 0,976 | 0,999 | 0,976 |
| PVE_3 | 0,012 | 0,004 | 0,003 | 0,010 |
| PVE_2 | 0,051 | 0,063 | 0,065 | 0,068 |
| PVE | 0,529 | 0,594 | 0,605 | 0,681 |
| MAE | 0,430 | 0,375 | 0,369 | 0,310 |
| $RMSE$ | 0,591 | 0,506 | 0,498 | 0,427 |
| $CV(RMSE)$ | 15,73 | 13,51 | 13,32 | 11,43 |
| $NMBE$ | 0,02 | 0,02 | 0,03 | 0,05 |
| Prop. Est. Elim | 0 | 0,102 | 0,117 | 0,602 |
| N° Est | 3243 | 2913 | 2863 | 1290 |

Invarianza de las medidas.

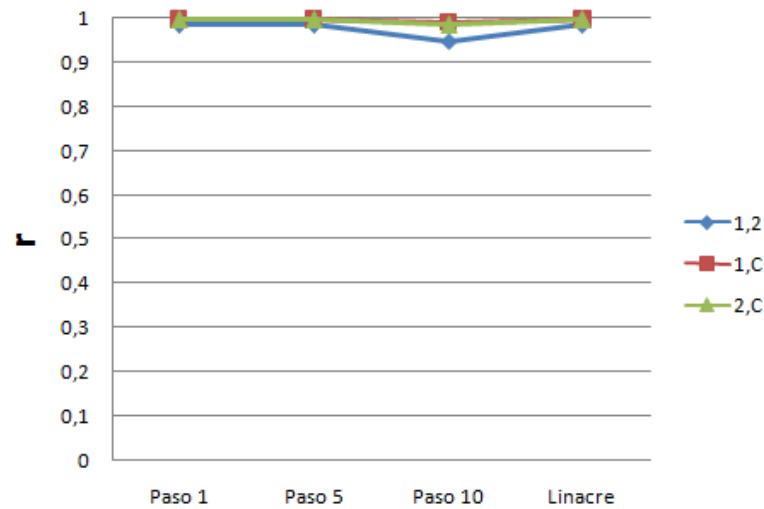
En relación a la invarianza de las medidas, si el algoritmo diagnóstico propuesto funciona adecuadamente, se espera que: (a) la invarianza de las medidas mejore o se mantenga similar a la invarianza inicial, a medida que se avanza en los pasos y etapas del algoritmo de diagnóstico; y (b), la invarianza final sea similar o mejor que la obtenida siguiendo estrictamente el algoritmo propuesto por Linacre.

En la Figura 5.8 se muestra la invarianza de las medidas de los ítems y docentes para los pasos 1, 5 y 10 (método alternativo propuesto) y la invarianza final obtenida siguiendo estrictamente el algoritmo sugerido por Linacre. Observar que, $r_{1,2}$ denota el coeficiente de correlación lineal entre las medidas obtenidas a partir de dos sub-muestras aleatorias sin elementos en común, mientras que $r_{1,C}$ y $r_{2,C}$ son los coeficientes de correlación lineal entre las medidas obtenidas a partir dichas sub-muestras con las medidas obtenidas a partir

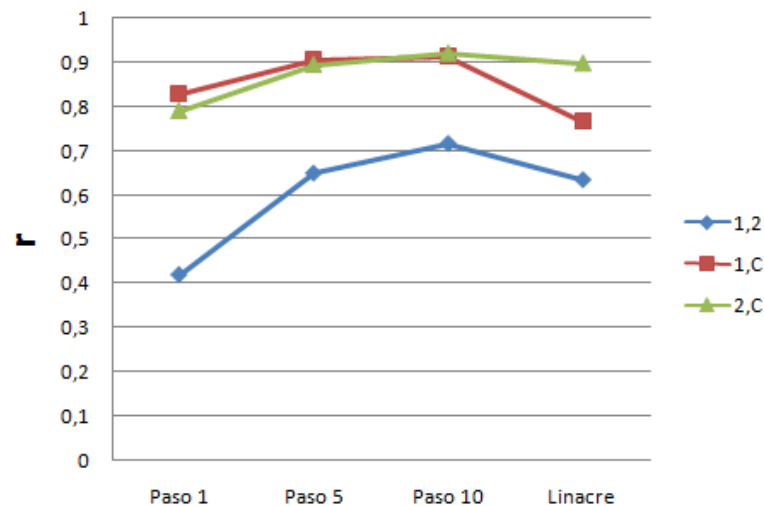
de la muestra completa, respectivamente. Se aprecia una notable invarianza de las medidas de los ítems para ambos algoritmos de diagnóstico. No obstante, la invarianza para el paso 5 del método alternativo ($r_{1,2} = 0,984$; $r_{1,C} = 0,996$; y $r_{2,C} = 0,996$) es un poco mayor que la obtenida siguiendo estrictamente el algoritmo sugerido por Linacre ($r_{1,2} = 0,950$; $r_{1,C} = 0,988$; y $r_{2,C} = 0,986$).

Adicionalmente, la Figura 5.8a sugiere una pequeña disminución en la invarianza entre el paso 5 y el paso 10. Esto podría estar reflejando el efecto de la eliminación de elementos que son importantes para mantener una buena conexión entre los datos, y por tanto, una disminución en la exactitud de las estimaciones de las medidas de los ítems. También, podría estar evidenciando una menor precisión en las estimaciones, debido a una disminución en el número de observaciones en las que se basan las estimaciones. Al corregir la atenuación de los coeficientes de correlación lineal, entre las medidas obtenidas, se aprecia que no hay diferencia en los valores entre el paso 5 y 10 para $r_{1,C}$ y $r_{2,C}$, pero, si para $r_{1,2}$, valor que se espera sea más afectado por problemas de conexión o solapamiento entre elementos, que pueden producir una menor exactitud en las estimaciones.

En la Figura 5.8b se aprecia que la invarianza de las medidas de los docentes respecto a los estudiantes es menor que para las medidas de los ítems. El patrón observado en la Figura 5.8b sugiere una menor exactitud en las estimaciones derivada de una menor conexión entre los elementos de la faceta ítems respecto de la faceta docentes. Esto se evidencia claramente en la gran diferencia observada entre $r_{1,2}$ y los valores de $r_{1,C}$ y $r_{2,C}$ para los dos algoritmos de diagnóstico comparados. También se aprecia, que el algoritmo propuesto se comporta mejor que el algoritmo sugerido por Linacre. Nuevamente, la invarianza para el paso 5 del método alternativo ($r_{1,2} = 0,649$; $r_{1,C} = 0,906$; y $r_{2,C} = 0,895$) es mayor que la obtenida siguiendo estrictamente el algoritmo sugerido por Linacre ($r_{1,2} = 0,634$; $r_{1,C} = 0,767$; y $r_{2,C} = 0,868$).



(a) Invarianza en las medidas de los ítems respecto a los estudiantes.



(b) Invarianza en las medidas de los docentes respecto a los estudiantes.

Figura 5.8: Comparación de la invarianza de las medidas respecto a los estudiantes según pasos de diagnóstico (1,5 y 10) y métodos (método de Linacre y método alternativo).

5.7. Sumario y Conclusiones.

La conexión en la faceta docentes resultó mayor que la observada en la faceta estudiantes. La red de evaluación encontrada resultó ser una red de

evaluación conectada incompleta e incluyó conjuntos de elementos que son claves para la conexión de la red. Los criterios de comparación considerados sugieren que el método propuesto funciona mejor.

A partir de la evaluación empírica de la estabilización de las medidas, ajuste global e invarianza para el método diagnóstico propuesto y su comparación con la aplicación estricta de las orientaciones dadas por Linacre, se considera una aproximación razonable aplicar el algoritmo propuesto con un criterio de parada menos estricto, esto es un $\% \Delta M_n$ cercano al $\%5$. Este criterio de parada coincide aproximadamente con aplicar el criterio diagnóstico durante los primeros 5 pasos.

Por tanto, una aproximación útil y practicable resulta ser aplicar el criterio diagnóstico hasta el paso 5 y analizar como se comporta las estimaciones de las medidas, el ajuste global del modelo, de los elementos de cada faceta y la invarianza. Las evaluaciones empíricas realizadas previamente sugieren que este procedimiento funciona mejor, que el algoritmo propuesto por Linacre, que presupone un mejor solapamiento o conexión entre los elementos de las facetas, que el observado en este caso.

Respecto al problema central abordado en esta tesis, el método diagnóstico propuesto resulta mucho más adecuado debido a que permite considerar el proceso de respuesta de un alto porcentaje de los evaluadores (estudiantes) en la definición de la escala de medida (90 % en comparación con un 40 %).

Capítulo 6

Validez de un cuestionario de evaluación estudiantil de la calidad en la enseñanza: análisis de múltiples facetas de Rasch.

6.1. Presentación.

Un primer objetivo de la investigación reportada en este capítulo es evaluar la validez de las inferencias que pueden realizarse a partir de un cuestionario de evaluación de la calidad en la enseñanza y estudiar sus propiedades psicométricas. Un segundo objetivo es evaluar la validez de algunas predicciones teóricas derivadas de la hipótesis de investigación central propuesta en esta tesis.

La hipótesis fundamental en nuestra investigación es que existen diferencias individuales en cómo los estudiantes evalúan la enseñanza que están relacionadas con variables que median la interpretación cognitiva de las características de la enseñanza o bien explican diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes. La principal predicción realizada es que la faceta estudiantes debería tener un menor ajuste al modelo que las facetas docentes e ítems. A partir de esta hipótesis se predice que el menor ajuste esperado de la faceta estudiantes estaría relacionada con cambios en la severidad del estudiante que serían resultado de la interacción entre características del estudiante y las características de la enseñanza.

El cuestionario estudiado consistió en 49 ítems organizados en 9 sub-escalas. Estas sub-escalas fueron desarrolladas siguiendo procedimientos requeridos para obtener instrumentos con una adecuada validez de contenido y aparente. La muestra de datos de calibración correspondió a 3243 estudiantes que evaluaron en conjunto 98 docentes de 10 actividades de enseñanza y siete disciplinas diferentes. Estos datos corresponden a estudiantes que respondieron las escalas consideradas, durante dos semestres consecutivos, en los primeros tres años de la carrera de médico de la Facultad de Medicina de la Universidad de la República durante el año 2015. Las respuestas de los estudiantes a cada una de las sub-escalas fueron analizadas por separado mediante un modelo de múltiples facetas de Rasch con tres facetas de medición (docentes, estudiantes e ítems) y una faceta “dummy” (curso).

Se encontró un buen ajuste global entre el modelo y los datos. En relación al ajuste de cada uno de los elementos de las facetas se evidenció un buen ajuste en todos los ítems y todos los docentes. Se destaca una notable invarianza de las medidas de calidad de la enseñanza de los docentes respecto a los ítems.

Los resultados muestran también una excelente confiabilidad de separación para cada una de las dimensiones de la calidad en la enseñanza y para la dificultad de los ítems. La invarianza de las medidas de la calidad de la enseñanza de los docentes respecto a los estudiantes resultó menor que las otras invarianzas consideradas.

En relación al ajuste de la faceta estudiantes se encontró un nivel de ajuste menor que para las otras facetas consideradas y se encontró un efecto de evaluador significativo que refleja diferencias en la severidad de los estudiantes al evaluar la enseñanza. Se encontraron también evidencias de funcionamiento diferencial de los evaluadores (estudiantes).

Los resultados encontrados aportan evidencias de la validez del cuestionario considerado y son consistentes con las predicciones realizadas a partir de nuestra hipótesis de investigación.

6.2. Metodología.

La metodología básica empleada en la presente investigación fue descrita en el capítulo 3 y en el capítulo 5 sección 5.2. En esta sección se describen aspectos específicos correspondientes a este capítulo.

6.2.1. Muestra de datos de calibración.

La muestra de datos de calibración considerada aquí consistió en las respuestas a las 9 sub-escalas del cuestionario de evaluación estudiantil de la calidad en la enseñanza de 3243 estudiantes, que evaluaron 98 docentes, en 10 actividades de enseñanza, durante dos semestres consecutivos, en la carrera de médico, durante el año 2015.

6.2.2. Instrumentos.

El instrumento considerado aquí consiste en las 9 sub-escalas de evaluación de la calidad de la enseñanza desarrollado para medir dimensiones relacionadas con las medidas por “Endeavor Instrument” (Frey y cols., 1975) y “Students’ Evaluation of Educational Quality” o SEEQ (Marsh, 1987).

6.2.3. Aproximación Analítica.

El modelo de múltiples facetas de Rasch básico utilizado aquí se describió en detalle en la sección 5.2 del capítulo 5 de esta tesis. Este modelo contiene tres facetas de medición (docentes, estudiantes e ítems) y una faceta “dummy” (curso). Con este modelo, se llevó a cabo el análisis de cada una de las 9 sub-escalas por separado. El procedimiento de diagnóstico utilizado fue descrito en los capítulos 3 y 5. Para estudiar las interacciones entre elementos de las distintas facetas se consideró el modelo básico más la adición de un término de interacción.

6.3. Resultados.

6.3.1. Ajuste global entre datos y el modelo.

Los resultados reportados en esta sección corresponde al ajuste final obtenido mediante el algoritmo interactivo discutido en la sección 5.4 aplicado al modelo descrito en la sección 5.2. Cada sub-escala fue analizada por separado. El número de estudiantes considerados en el paso final y la proporción de estudiantes eliminados se muestra para cada sub-escala. El número de estudiantes eliminados hasta el paso final del algoritmo diagnóstico osciló en un valor próximo a un 10 %.

Se encontró un buen ajuste global entre los datos y el modelo para todas las sub-escalas consideradas (Tabla 6.1). Los valores de *Infit* y *Outfit* para todas las escalas muestran un ajuste satisfactorio (Linacre, 2014). Estos valores incluso verifican el criterio más exigente propuesto por R. M. Smith y cols. (1995). Además, los valores obtenidos de *Infit* y *Outfit* muestran también que la aleatoriedad observada en los patrones de respuesta son consistentes con el modelo.

Tabla 6.1: Ajuste Global entre los datos y el modelo.

| | VALR | INTR | RELP | TRDF | CLOP | ENTC | APPR | TRPR | EVAL |
|----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 |
| <i>Infit</i> | 0,99 | 1,00 | 0,99 | 1,00 | 0,99 | 0,99 | 0,99 | 0,99 | 1,00 |
| <i>Outfit</i> | 0,98 | 0,98 | 0,97 | 0,98 | 0,97 | 0,97 | 0,98 | 0,94 | 0,99 |
| <i>Media</i> (Z_{Res}) | 0,00 | 0,00 | 0,02 | 0,00 | 0,01 | 0,01 | 0,00 | 0,01 | 0,00 |
| <i>SD</i> (Z_{Res}) | 0,98 | 0,98 | 0,94 | 0,99 | 0,96 | 0,96 | 0,97 | 0,95 | 0,99 |
| PRE_3 | 0,00 | 0,01 | 0,00 | 0,01 | 0,00 | 0,01 | 0,00 | 0,00 | 0,00 |
| PRE_2 | 0,06 | 0,06 | 0,05 | 0,06 | 0,06 | 0,05 | 0,06 | 0,07 | 0,05 |
| PVE | 0,59 | 0,56 | 0,64 | 0,57 | 0,62 | 0,58 | 0,58 | 0,66 | 0,53 |
| MAE | 0,38 | 0,40 | 0,34 | 0,33 | 0,36 | 0,40 | 0,37 | 0,30 | 0,48 |
| RMSE | 0,51 | 0,54 | 0,48 | 0,57 | 0,49 | 0,54 | 0,51 | 0,44 | 0,62 |
| CV(RMSE) | 13,51 | 14,48 | 12,22 | 17,75 | 12,92 | 14,25 | 13,52 | 12,18 | 19,25 |
| NMBE | 0,02 | 0,02 | 0,09 | 0,01 | 0,03 | 0,05 | 0,03 | 0,05 | 0,01 |
| Prop. Est. Elim. | 0,10 | 0,11 | 0,09 | 0,10 | 0,11 | 0,11 | 0,07 | 0,06 | 0,10 |
| N° Est | 2913 | 2874 | 2951 | 2907 | 2887 | 2889 | 3005 | 3051 | 2925 |

Adicionalmente, los valores obtenidos de la media ($Media(Z_{Res})$) y desvío estándar ($SD(Z_{Res})$), de los residuos estandarizados (Z_{Res}) para el modelo, son muy próximos a los valores esperados para un ajuste perfecto entre los datos y el modelo ($Media(Z_{Res}) = 0$ y $SD(Z_{Res}) = 1$). En forma similar las proporciones de residuos estandarizados absolutos extremos (PRE_3 y PRE_2) son muy próximos a los valores esperados para un ajuste perfecto ($PRE_3 = 0,00$ y $PRE_2 = 0,05$).

Por su parte, cuando se analiza el ajuste absoluto global entre los datos y el modelo enfatizando su exactitud y precisión en la predicción de los puntajes observados, se encuentra también un ajuste satisfactorio. Los valores de *MAE* mostrados en la Tabla 6.1, muestran que el error absoluto promedio, o diferencia absoluta promedio entre el puntaje observado y el predicho por el modelo, es relativamente pequeño. Los valores de *MAE* son próximos a un 10% del rango de variación del puntaje observado. Además, los valores de *MAE* son menores a la mitad de la diferencia mínima entre dos puntajes observados diferentes.

En forma similar, los valores de *RMSE* para el modelo son consistentes con un ajuste satisfactorio. *RMSE* puede ser interpretado como el desvío

estándar de los puntajes observados respecto a los valores predichos por el modelo. Los valores de $RMSE$ son relativamente bajos, como muestran los valores normalizados ($CV(RMSE)$) que expresan el valor de $RMSE$ como porcentaje respecto a la media de los puntajes observados.

Por otra parte, los valores de $NMBE$ (o sesgo promedio de los errores normalizado) muestran que el sesgo promedio de los valores predichos por el modelo es extraordinariamente bajo. Para todas las sub-escalas el valor de $NMBE$ es un orden de magnitud menor que 1 %.

Los valores de PVE en la Tabla 6.1 describen la proporción de la varianza de los puntajes observados explicada por el modelo. El valor de PVE es análogo al R^2 usado con modelos de regresión múltiple y puede considerarse una generalización, para variables politómicas, del *pseudo* - R^2 de Efron (1978). Como se aprecia en la Tabla 6.1 los valores de PVE oscilan entorno a un 60 % de la varianza de los puntajes observados. Por tanto, una proporción substancial de la varianza observada es explicada por el modelo.

En síntesis, los resultados del análisis del ajuste global entre los datos y el modelo muestran un ajuste satisfactorio. Esto es compatible con la posibilidad de que el modelo considerado pueda resultar útil como primera aproximación o base para modelizar el proceso de respuesta a estos cuestionarios. En otros términos, el resultado obtenido, sobre el ajuste global entre los datos y el modelo para cada una de las sub-escalas, sugiere que el modelo captura aspectos esenciales del proceso de respuesta que se quiere modelar. Adicionalmente, estos resultados son compatibles con la viabilidad de utilizar como base el modelo considerado para obtener medidas lineales aditivas invariantes de los elementos considerados en cada una de las facetas.

6.3.2. Análisis de las Facetas de Docentes, Estudiantes e Ítems.

En relación al ajuste de cada uno de los elementos de las facetas (estudiantes, docentes e ítems) se evidenció un buen ajuste en todos los ítems y todos los docentes considerados (véase Tablas 6.2, 6.3 y 6.5). Los resultados muestran también una excelente confiabilidad de separación (SR , o confiabilidad de Rasch) para cada una de las dimensiones de la calidad en la enseñanza consideradas y para la dificultad de los ítems (Tablas 6.4 y 6.6). En relación al ajuste de la faceta estudiantes se encontró un nivel de ajuste menor que para

las otras facetas consideradas (Tabla 6.8). En relación a la faceta estudiantes se encontró además un efecto de evaluador significativo que refleja diferencias en la severidad de los estudiantes al evaluar la enseñanza. Estos resultados son consistentes con la predicción teórica realizada a partir de nuestra hipótesis de investigación. Como se predijo, se encontró un mejor ajuste en la faceta ítems y faceta docentes, que en la faceta estudiantes.

Faceta Ítems.

En la Tabla 6.2 se muestran las medidas obtenidas para cada ítem, el error estándar de medición y estadísticos de ajuste. Los valores de Infit y Outfit obtenidos para cada ítems indican un alto nivel de ajuste al modelo. Los valores de correlación punto medida (r_{pm}) mostraron consistencia entre las medidas de cada ítem y el constructo medido por su escala, para todos los ítems y sub-escalas consideradas. Adicionalmente, los coeficientes de discriminación (*Discrim*) presentaron valores en general cercanos al valor 1,00 esperado si el ajuste entre los datos y el modelo es perfecto.

Los valores presentados en la Tabla 6.2 son resumidos para cada sub-escala en la Tabla 6.3, para mostrar la distribución de las medidas y estadísticos de ajuste de los ítems para cada sub-escala. En todos los casos los valores de Infit y Outfit medios son muy cercanos al valor 1 (ajuste perfecto) con un desvío estándar muy pequeño. En forma similar el valor de la media de r_{pm} indica un alto nivel de consistencia entre las medidas de los ítems y el constructo medido por la escala, también con un valor del desvío estándar de r_{pm} muy pequeño.

El análisis de la confiabilidad de las medidas de los ítems se presenta en la Tabla 6.4. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 ítems tienen diferente dificultad, para cada una de las sub-escalas consideradas.

En la Tabla 6.4 se muestra los valores de la razón de separación (G) para cada escala. G describe la amplitud de variación de las medidas de dificultad (SD_A) tomando como unidad el error estándar de medición ($RMSE$). Los valores de G muestran que en general las medidas de dificultad de los ítems de las sub-escalas son notablemente confiables.

Esta alta confiabilidad en las mediciones se refleja en el número de estratos estadísticamente discernibles (H) que podrían definirse para las dificultades de los ítems. Por ejemplo, si bien la escala EVAL(E7) contiene solo siete ítems, con

el nivel de confiabilidad de las mediciones encontrado, se podrían definir cerca de 43 estratos de dificultad. Se encontró también un alto valor de confiabilidad de separación (SR), consistente con los valores obtenidos de los otros índices de confiabilidad.

Si bien en general la confiabilidad de las mediciones resultó elevada, existen variaciones en la confiabilidad entre escalas. En particular, las medidas de los ítems obtenidas con la sub-escala TRPR(E8) presentaron una mucho menor confiabilidad que el resto, mientras que para la sub-escala EVAL(E9) se encontró un valor mucho más elevado de la confiabilidad de las medidas de las dificultades, que para las demás sub-escalas.



Tabla 6.2: Ítems : Medidas y estadísticos de ajuste por ítem

| Items | Medida | S.E. | Infit | Outfit | r_{pm} | Discrim |
|-------|--------|------|-------|--------|----------|---------|
| VALR1 | 0,24 | 0,02 | 1,04 | 1,04 | 0,77 | 0,95 |
| VALR2 | -0,37 | 0,02 | 0,94 | 0,91 | 0,79 | 1,06 |
| VALR3 | 0,04 | 0,02 | 0,95 | 0,93 | 0,79 | 1,05 |
| VALR4 | -0,09 | 0,02 | 1,04 | 1,04 | 0,76 | 0,95 |
| VALR5 | 0,18 | 0,02 | 1,00 | 0,98 | 0,78 | 0,99 |
| INTR1 | 0,29 | 0,02 | 1,02 | 1,02 | 0,76 | 0,98 |
| INTR2 | -0,04 | 0,02 | 0,93 | 0,91 | 0,77 | 1,07 |
| INTR3 | -0,15 | 0,02 | 0,92 | 0,90 | 0,77 | 1,08 |
| INTR4 | 0,44 | 0,02 | 1,14 | 1,13 | 0,73 | 0,85 |
| INTR5 | -0,11 | 0,02 | 0,96 | 0,94 | 0,76 | 1,04 |
| INTR6 | -0,44 | 0,02 | 1,02 | 0,99 | 0,74 | 0,99 |
| RELP1 | 0,22 | 0,02 | 1,03 | 1,01 | 0,77 | 0,93 |
| RELP2 | -0,5 | 0,02 | 0,95 | 0,91 | 0,77 | 1,00 |
| RELP3 | 0,00 | 0,02 | 0,93 | 0,89 | 0,78 | 1,03 |
| RELP4 | 0,38 | 0,02 | 1,12 | 1,14 | 0,76 | 0,83 |
| RELP5 | -0,10 | 0,02 | 0,94 | 0,89 | 0,78 | 1,02 |
| TRDF1 | -0,54 | 0,02 | 1,05 | 1,04 | 0,75 | 0,94 |
| TRDF2 | 0,21 | 0,02 | 0,93 | 0,91 | 0,77 | 1,06 |
| TRDF3 | 0,24 | 0,02 | 1,04 | 1,02 | 0,75 | 0,97 |
| TRDF4 | 0,09 | 0,02 | 0,98 | 0,97 | 0,77 | 1,02 |
| CLOR1 | 0,14 | 0,02 | 0,97 | 0,96 | 0,79 | 1,02 |
| CLOR2 | 0,15 | 0,02 | 0,97 | 0,95 | 0,79 | 1,02 |
| CLOR3 | 0,02 | 0,02 | 0,98 | 0,96 | 0,79 | 1,01 |
| CLOR4 | -0,16 | 0,02 | 0,98 | 0,96 | 0,78 | 1,01 |
| CLOR5 | -0,15 | 0,02 | 0,98 | 0,95 | 0,78 | 1,01 |
| CLOR6 | 0,09 | 0,02 | 0,98 | 0,96 | 0,79 | 1,01 |
| CLOR7 | 0,09 | 0,02 | 1,07 | 1,05 | 0,78 | 0,93 |
| CLOR8 | -0,17 | 0,02 | 1,01 | 0,98 | 0,78 | 0,99 |
| ENTC1 | -0,18 | 0,02 | 0,90 | 0,89 | 0,75 | 1,10 |
| ENTC2 | -0,1 | 0,02 | 0,91 | 0,88 | 0,75 | 1,09 |
| ENTC3 | 0,29 | 0,02 | 1,06 | 1,05 | 0,73 | 0,92 |
| ENTC4 | 0,26 | 0,02 | 1,14 | 1,12 | 0,72 | 0,84 |
| ENTC5 | -0,26 | 0,02 | 0,97 | 0,94 | 0,73 | 1,02 |
| APPR1 | -0,04 | 0,02 | 1,00 | 0,99 | 0,78 | 1,00 |
| APPR2 | 0,15 | 0,02 | 1,01 | 1,01 | 0,77 | 0,98 |
| APPR3 | 0,13 | 0,02 | 1,02 | 1,02 | 0,77 | 0,97 |
| APPR4 | 0,00 | 0,02 | 0,94 | 0,92 | 0,79 | 1,06 |
| APPR5 | -0,24 | 0,02 | 1,00 | 0,98 | 0,77 | 0,99 |
| TRPR1 | 0,00 | 0,03 | 0,95 | 0,91 | 0,84 | 1,04 |
| TRPR2 | 0,09 | 0,03 | 0,93 | 0,89 | 0,85 | 1,05 |
| TRPR3 | 0,03 | 0,02 | 1,13 | 1,08 | 0,81 | 0,87 |
| TRPR4 | -0,11 | 0,02 | 0,95 | 0,89 | 0,84 | 1,04 |
| EVAL1 | -0,27 | 0,02 | 0,86 | 0,85 | 0,75 | 1,13 |
| EVAL2 | 0,34 | 0,02 | 0,87 | 0,87 | 0,76 | 1,13 |
| EVAL3 | -0,02 | 0,02 | 0,90 | 0,89 | 0,74 | 1,09 |
| EVAL4 | 0,83 | 0,02 | 1,02 | 1,03 | 0,73 | 0,98 |
| EVAL5 | -0,42 | 0,02 | 0,96 | 0,95 | 0,71 | 1,04 |
| EVAL6 | -1,01 | 0,02 | 1,22 | 1,19 | 0,62 | 0,77 |
| EVAL7 | 0,55 | 0,02 | 1,15 | 1,17 | 0,68 | 0,84 |

Tabla 6.3: Estadísticos descriptivos para la faceta ítems para cada sub-escala

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| VALR E1 | Media | 0,00 | 0,02 | 0,99 | 0,98 | 0,78 |
| | SD | 0,20 | 0,00 | 0,04 | 0,05 | 0,01 |
| | Q1 | -0,07 | 0,02 | 0,96 | 0,94 | 0,77 |
| | Q3 | 0,18 | 0,02 | 1,04 | 1,04 | 0,79 |
| INTR E2 | Media | 0,00 | 0,02 | 1,00 | 0,98 | 0,76 |
| | SD | 0,29 | 0,00 | 0,07 | 0,08 | 0,02 |
| | Q1 | -0,14 | 0,02 | 0,94 | 0,92 | 0,75 |
| | Q3 | -0,08 | 0,02 | 0,99 | 0,97 | 0,76 |
| RELP E3 | Media | 0,00 | 0,02 | 0,99 | 0,97 | 0,77 |
| | SD | 0,30 | 0,00 | 0,07 | 0,10 | 0,01 |
| | Q1 | -0,10 | 0,02 | 0,94 | 0,89 | 0,77 |
| | Q3 | 0,22 | 0,02 | 1,03 | 1,01 | 0,78 |
| TRDF E4 | Media | 0,00 | 0,02 | 1,00 | 0,99 | 0,76 |
| | SD | 0,32 | 0,00 | 0,05 | 0,05 | 0,01 |
| | Q1 | -0,07 | 0,02 | 0,97 | 0,96 | 0,75 |
| | Q3 | 0,22 | 0,02 | 1,04 | 1,03 | 0,77 |
| CLOR E5 | Media | 0,00 | 0,02 | 0,99 | 0,97 | 0,79 |
| | SD | 0,13 | 0,00 | 0,03 | 0,03 | 0,01 |
| | Q1 | -0,15 | 0,02 | 0,98 | 0,96 | 0,78 |
| | Q3 | 0,06 | 0,02 | 0,98 | 0,96 | 0,79 |
| ENTC E6 | Media | 0,00 | 0,02 | 1,00 | 0,98 | 0,74 |
| | SD | 0,23 | 0,00 | 0,09 | 0,09 | 0,01 |
| | Q1 | -0,18 | 0,02 | 0,91 | 0,89 | 0,73 |
| | Q3 | 0,26 | 0,02 | 1,06 | 1,05 | 0,75 |
| APPR E7 | Media | 0,00 | 0,02 | 0,99 | 0,98 | 0,78 |
| | SD | 0,14 | 0,00 | 0,03 | 0,03 | 0,01 |
| | Q1 | -0,04 | 0,02 | 1,00 | 0,98 | 0,77 |
| | Q3 | 0,13 | 0,02 | 1,01 | 1,01 | 0,78 |
| TRPR E8 | Media | 0,00 | 0,03 | 0,99 | 0,94 | 0,84 |
| | SD | 0,07 | 0,01 | 0,08 | 0,08 | 0,02 |
| | Q1 | -0,03 | 0,02 | 0,95 | 0,89 | 0,83 |
| | Q3 | 0,05 | 0,03 | 1,00 | 0,95 | 0,84 |
| EVAL E9 | Media | 0,00 | 0,02 | 1,00 | 0,99 | 0,71 |
| | SD | 0,58 | 0,00 | 0,13 | 0,13 | 0,05 |
| | Q1 | -0,35 | 0,02 | 0,89 | 0,88 | 0,70 |
| | Q3 | 0,45 | 0,02 | 1,09 | 1,10 | 0,75 |

Tabla 6.4: Confiabilidad de Separación para la faceta Ítems para cada sub-escala

| | <i>SR</i> | <i>G</i> | <i>H</i> | Prueba Efectos Fijos (todos iguales) valor p |
|----------|-----------|----------|----------|----------------------------------------------------------|
| VALR(E1) | 0,99 | 9,85 | 13,46 | 0,00 |
| INTR(E2) | 0,99 | 13,87 | 18,83 | 0,00 |
| RELP(E3) | 0,99 | 12,75 | 17,33 | 0,00 |
| TRDF(E4) | 0,99 | 15,93 | 21,58 | 0,00 |
| CLOR(E5) | 0,97 | 5,63 | 7,84 | 0,00 |
| ENTC(E6) | 0,99 | 10,91 | 14,87 | 0,00 |
| APPR(E7) | 0,98 | 6,33 | 8,77 | 0,00 |
| TRPR(E8) | 0,88 | 2,66 | 3,88 | 0,00 |
| EVAL(E9) | 1,00 | 31,91 | 42,89 | 0,00 |

Faceta Docentes.

En la Tabla 6.5 se muestran las distribuciones de las medidas y estadísticos de ajuste obtenidas para los docentes, para cada sub-escala. Las distribuciones que se encontraron de los valores de Infit y Outfit de los docentes, para todas las sub-escalas, indican un buen nivel de ajuste. No obstante, el ajuste encontrado en los docentes es menor que el ajuste encontrado en los ítems. Esto queda en evidencia al comparar los valores medios y desvíos estándar encontrados para los índices de ajuste Infit y Outfit, entre la faceta docentes e ítems.

Por otra parte, el valor de la media y desvío estándar de r_{pm} , mostrado en la Tabla 6.5 indica un alto nivel de consistencia entre las medidas de los docentes y el constructo medido por la escala. Por otra parte, el nivel de consistencia encontrado entre las medidas de los docentes y las medidas de las observaciones para cada sub-escala es menor que la observada para los ítems. Por lo tanto, si bien el nivel de acuerdo entre el modelo y los datos, para la faceta docentes es bueno, es menor que el observado para los ítems.

El análisis de la confiabilidad de las medidas de los docentes se presenta en la Tabla 6.6. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 docentes tienen diferente calidad en la enseñanza, para cada una de las sub-escalas consideradas.

Los valores de G mostraron que en general las medidas de calidad o efectividad de los docentes, obtenidas a partir de cada sub-escala, presentaron una elevada confiabilidad (G entre 5,75 (sub-escala EVAL E9) y 16,17 (sub-escala RELP E3)). No obstante, se observan variaciones importantes en los valores de G entre las sub-escalas. Esta elevada confiabilidad, se apreció también, cuando se consideró el índice SR , o confiabilidad de Rasch, análogo de la confiabilidad de un test, en la teoría clásica de los test. Los valores de SR oscilaron entre 0,97 (sub-escala EVAL E9) y 1,00 (sub-escala RELP E3).

Esta alta confiabilidad en las mediciones de la calidad o efectividad se reflejó en el número de estratos estadísticamente discernibles (H) que podrían definirse para las medidas de los docentes. El valor de H para las medidas de los docentes osciló entre 8 (sub-escala EVAL E9) y cerca de 22 (sub-escala RELP E3).

Tabla 6.5: Estadísticos descriptivos para la faceta docentes para cada sub-escala

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| VALR E1 | Media | 2,29 | 0,16 | 1,11 | 1,11 | 0,71 |
| | SD | 1,38 | 0,08 | 0,28 | 0,30 | 0,16 |
| | Q1 | 1,44 | 0,12 | 0,96 | 0,93 | 0,69 |
| | Q3 | 2,77 | 0,19 | 1,29 | 1,31 | 0,78 |
| INTR E2 | Media | 2,18 | 0,14 | 1,13 | 1,12 | 0,70 |
| | SD | 1,21 | 0,07 | 0,31 | 0,32 | 0,16 |
| | Q1 | 1,45 | 0,10 | 0,96 | 0,95 | 0,67 |
| | Q3 | 2,84 | 0,17 | 1,36 | 1,35 | 0,76 |
| RELP E3 | Media | 3,71 | 0,18 | 1,09 | 1,08 | 0,76 |
| | SD | 3,27 | 0,09 | 0,35 | 0,39 | 0,09 |
| | Q1 | 2,31 | 0,11 | 0,86 | 0,83 | 0,72 |
| | Q3 | 4,23 | 0,22 | 1,29 | 1,33 | 0,82 |
| TRDF E4 | Media | 1,19 | 0,16 | 1,20 | 1,20 | 0,72 |
| | SD | 0,96 | 0,08 | 0,31 | 0,32 | 0,08 |
| | Q1 | 0,60 | 0,11 | 1,02 | 0,98 | 0,67 |
| | Q3 | 1,74 | 0,19 | 1,37 | 1,39 | 0,77 |
| CLOR E5 | Media | 3,05 | 0,13 | 1,09 | 1,09 | 0,71 |
| | SD | 2,40 | 0,07 | 0,33 | 0,36 | 0,19 |
| | Q1 | 1,81 | 0,09 | 0,87 | 0,83 | 0,67 |
| | Q3 | 3,82 | 0,16 | 1,28 | 1,32 | 0,79 |
| ENTC E6 | Media | 2,56 | 0,16 | 1,08 | 1,07 | 0,70 |
| | SD | 2,57 | 0,08 | 0,32 | 0,34 | 0,11 |
| | Q1 | 1,44 | 0,10 | 0,88 | 0,86 | 0,65 |
| | Q3 | 3,18 | 0,19 | 1,27 | 1,28 | 0,76 |
| APPR E7 | Media | 2,63 | 0,16 | 1,13 | 1,14 | 0,68 |
| | SD | 1,32 | 0,08 | 0,31 | 0,33 | 0,22 |
| | Q1 | 1,77 | 0,11 | 0,90 | 0,90 | 0,65 |
| | Q3 | 3,67 | 0,19 | 1,32 | 1,32 | 0,78 |
| TRPR E8 | Media | 2,67 | 0,20 | 1,19 | 1,15 | 0,78 |
| | SD | 1,51 | 0,10 | 0,42 | 0,42 | 0,10 |
| | Q1 | 1,89 | 0,14 | 0,84 | 0,80 | 0,73 |
| | Q3 | 3,23 | 0,25 | 1,46 | 1,39 | 0,85 |
| EVAL E9 | Media | 0,77 | 0,11 | 1,06 | 1,06 | 0,68 |
| | SD | 0,72 | 0,05 | 0,25 | 0,25 | 0,10 |
| | Q1 | 0,45 | 0,08 | 0,88 | 0,88 | 0,64 |
| | Q3 | 1,09 | 0,13 | 1,25 | 1,25 | 0,73 |

Tabla 6.6: Confiabilidad de Separación para la faceta Docentes para cada sub-escala

| | <i>SR</i> | <i>G</i> | <i>H</i> | Prueba Efectos Fijos (todos iguales) valor p |
|----------|-----------|----------|----------|----------------------------------------------|
| VALR(E1) | 0,98 | 7,54 | 10,39 | 0,00 |
| INTR(E2) | 0,98 | 7,53 | 10,38 | 0,00 |
| RELP(E3) | 1,00 | 16,17 | 21,90 | 0,00 |
| TRDF(E4) | 1,00 | 5,16 | 7,21 | 0,00 |
| CLOR(E5) | 1,00 | 15,83 | 21,44 | 0,00 |
| ENTC(E6) | 1,00 | 14,51 | 19,68 | 0,00 |
| APPR(E7) | 0,98 | 7,20 | 9,93 | 0,00 |
| TRPR(E8) | 0,98 | 6,69 | 9,25 | 0,00 |
| EVAL(E9) | 0,97 | 5,75 | 8,00 | 0,00 |

Faceta Estudiantes.

En la Tabla 6.7 se muestran las distribuciones de las medidas y estadísticos de ajuste obtenidas para los estudiantes, para cada sub-escala. Las distribuciones que se encontraron de los valores de Infit y Outfit de los estudiantes, para todas las sub-escalas, indican un nivel de ajuste aceptable. No obstante, el ajuste encontrado en los estudiantes es menor que el ajuste encontrado para los ítems y docentes. Esto queda en evidencia al comparar los valores medios y desvíos estándar encontrados para los índices de ajuste Infit y Outfit, entre la faceta estudiantes, y las facetas docentes e ítems.

Por otra parte, la distribución de los valores de r_{pm} para cada una de las sub-escalas, mostrada en la Tabla 6.7 indica que en general se encontró un nivel de consistencia aceptable entre las medidas de los estudiantes y el constructo medido por la sub-escala. Para todas las sub-escalas se encuentra que al menos un 75% de los valores encontrados de r_{pm} son no negativos. Por otra parte, el nivel de consistencia encontrado entre las medidas de los estudiantes y las medidas de las observaciones para cada sub-escala es menor que la observada para los ítems y docentes. Esto se aprecia claramente al comparar la distribución de r_{pm} para los estudiantes con la distribución de r_{pm} encontrada en las otras facetas.

El análisis de la confiabilidad de las medidas de los estudiantes se presenta en la Tabla 6.8. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 estudiantes tienen diferente severidad al evaluar la enseñanza, para cada una de las sub-escalas consideradas.

Los valores de G mostraron que en general las medidas de severidad de los estudiantes, obtenidas a partir de cada sub-escala, presentaron una confiabilidad entre buena y elevada (G entre 3,08 (sub-escala ENTC E6) y 3,75 (sub-escala CLOR E5)), pero, claramente menores a las encontradas para las otras facetas. En este caso, por otra parte, no se observaron variaciones importantes en los valores de G entre las sub-escalas.

Los niveles de confiabilidad que fueron observados en esta faceta, se aprecian también, cuando se consideró el índice SR . Los valores de SR oscilaron entre 0,90 (sub-escala ENTC E6) y 0,93 (sub-escala CLOR E5). Observar que el valor de SR no tiene la misma interpretación que la confiabilidad entre evaluadores. SR describe la confiabilidad de separación de las medidas de ser-

verdad, esto significa cuán reproducible es la separación de los evaluadores por sus medidas de severidad. Por el contrario, la confiabilidad entre evaluadores, esta relacionada con cuán reproduciblemente las medidas de severidad son similares entre sí.

Una estadística de Rasch de la confiabilidad entre evaluadores, análoga a la utilizada en teoría clásica de los test, es el complemento de los valores de SR (Rasch-IRR o confiabilidad de Rasch entre evaluadores = $1 - SR$, véase [Linacre \(2015\)](#)). Por tanto, la confiabilidad entre evaluadores osciló entre 0,07 y 0,10, lo que representa una confiabilidad entre evaluadores muy baja.

Con los niveles de confiabilidad de las medidas de severidad que fueron encontrados en las escalas se estimó que el número de estratos estadísticamente discernibles de severidad de los estudiantes osciló entre 4 ($H = 4,17$, escala TRDF E4) y 5 ($H = 5,33$, escala CLOR E5).

Los resultados encontrados para las medidas de la severidad de los estudiantes claramente mostraron un efecto del evaluador en las evaluaciones, esto es, diferencias estadísticamente discernibles entre las severidades de los estudiantes.

Tabla 6.7: Estadísticos descriptivos para la faceta estudiantes para cada sub-escala

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| VALR E1 | Media | 0,00 | 0,70 | 0,98 | 0,98 | 0,26 |
| | SD | 2,53 | 0,33 | 0,75 | 0,76 | 0,40 |
| | Q1 | -1,50 | 0,42 | 0,51 | 0,48 | 0,00 |
| | Q3 | 1,65 | 0,85 | 1,30 | 1,33 | 0,57 |
| INTR E2 | Media | 0,00 | 0,62 | 0,99 | 0,99 | 0,28 |
| | SD | 2,41 | 0,34 | 0,75 | 0,75 | 0,38 |
| | Q1 | -1,39 | 0,36 | 0,46 | 0,45 | 0,00 |
| | Q3 | 1,49 | 0,72 | 1,40 | 1,42 | 0,60 |
| RELP E3 | Media | 0,00 | 0,85 | 0,95 | 0,95 | 0,27 |
| | SD | 3,41 | 0,52 | 0,76 | 0,80 | 0,39 |
| | Q1 | -1,96 | 0,47 | 0,46 | 0,43 | 0,00 |
| | Q3 | 1,72 | 0,97 | 1,14 | 1,16 | 0,63 |
| TRDF E4 | Media | 0,00 | 0,69 | 0,98 | 0,97 | 0,26 |
| | SD | 2,29 | 0,31 | 0,78 | 0,78 | 0,45 |
| | Q1 | -1,35 | 0,44 | 0,43 | 0,42 | 0,00 |
| | Q3 | 1,32 | 0,87 | 1,40 | 1,41 | 0,63 |
| CLOR E5 | Media | 0,00 | 0,64 | 0,95 | 0,95 | 0,19 |
| | SD | 2,98 | 0,43 | 0,73 | 0,75 | 0,41 |
| | Q1 | -1,77 | 0,35 | 0,49 | 0,45 | 0,00 |
| | Q3 | -0,06 | 0,57 | 0,91 | 0,88 | 0,10 |
| ENTC E6 | Media | 0,00 | 0,72 | 0,97 | 0,96 | 0,26 |
| | SD | 2,73 | 0,45 | 0,73 | 0,74 | 0,41 |
| | Q1 | -1,60 | 0,40 | 0,47 | 0,46 | 0,00 |
| | Q3 | 1,38 | 0,81 | 1,27 | 1,28 | 0,59 |
| APPR E7 | Media | 0,00 | 0,72 | 0,95 | 0,95 | 0,23 |
| | SD | 2,65 | 0,37 | 0,78 | 0,79 | 0,42 |
| | Q1 | -1,52 | 0,42 | 0,44 | 0,43 | 0,00 |
| | Q3 | 1,57 | 0,83 | 1,26 | 1,28 | 0,56 |
| TRPR E8 | Media | 0,00 | 0,90 | 0,91 | 0,91 | 0,12 |
| | SD | 3,42 | 0,40 | 0,98 | 0,99 | 0,47 |
| | Q1 | -2,22 | 0,55 | 0,03 | 0,03 | 0,00 |
| | Q3 | 2,25 | 1,25 | 1,00 | 1,02 | 0,44 |
| EVAL E9 | Media | 0,00 | 0,49 | 0,97 | 0,97 | 0,31 |
| | SD | 1,88 | 0,29 | 0,69 | 0,69 | 0,35 |
| | Q1 | -0,92 | 0,30 | 0,43 | 0,44 | 0,00 |
| | Q3 | 0,09 | 0,45 | 0,86 | 0,86 | 0,34 |

Tabla 6.8: Confiabilidad de Separación para la faceta Estudiantes para cada sub-escala

| | <i>SR</i> | <i>G</i> | <i>H</i> | Prueba Efectos Fijos (todos iguales) valor p |
|----------|-----------|----------|----------|----------------------------------------------------------|
| VALR(E1) | 0,91 | 3,11 | 4,47 | 0,00 |
| INTR(E2) | 0,91 | 3,25 | 4,67 | 0,00 |
| RELP(E3) | 0,92 | 3,28 | 4,71 | 0,00 |
| TRDF(E4) | 0,92 | 2,88 | 4,17 | 0,00 |
| CLOR(E5) | 0,93 | 3,75 | 5,33 | 0,00 |
| ENTC(E6) | 0,90 | 3,08 | 4,44 | 0,00 |
| APPR(E7) | 0,91 | 3,12 | 4,50 | 0,00 |
| TRPR(E8) | 0,92 | 3,32 | 4,76 | 0,00 |
| EVAL(E9) | 0,91 | 3,14 | 4,52 | 0,00 |

6.3.3. Invarianza.

Si existe un alto nivel de invarianza, se espera una correlación lineal elevada entre las medidas obtenidas para sub-muestras aleatorias complementarias, así como también correlaciones lineales elevadas entre cada sub-muestra y la muestra completa. En la Tabla 6.9 se muestran los resultados del análisis de invarianza de las medidas de elementos de las facetas ítems y docentes, respecto a los estudiantes, y de estudiantes y docentes, respecto a la faceta ítems.

Tabla 6.9: Análisis de la Invarianza de las Medidas para cada sub-escala

| Sub-escala | Correlación entre muestras | Invarianza Respecto a Estudiantes | | Invarianza Respecto a Ítems | |
|------------|----------------------------|-----------------------------------|---------------|-----------------------------|---------------|
| | | Ítems | Docentes | Estudiantes | Docentes |
| VALR (E1) | $r(1, 2)$ | 0,984 (1,000) | 0,649 (0,669) | 0,869 (1,000) | 0,941 (1,000) |
| | $r(1, C)$ | 0,996 (1,000) | 0,906 (0,929) | 0,948 (1,000) | 0,980 (1,000) |
| | $r(2, C)$ | 0,996 (1,000) | 0,895 (0,918) | 0,978 (1,000) | 0,990 (1,000) |
| INTR (E2) | $r(1, 2)$ | 0,987 (0,997) | 0,464 (0,478) | 0,874 (1,000) | 0,924 (0,948) |
| | $r(1, C)$ | 0,997 (1,000) | 0,870 (0,888) | 0,959 (1,000) | 0,976 (1,000) |
| | $r(2, C)$ | 0,997 (1,000) | 0,782 (0,798) | 0,971 (1,000) | 0,985 (1,000) |
| RELP (E3) | $r(1, 2)$ | 0,975 (0,985) | 0,843 (0,852) | 0,941 (1,000) | 0,993 (1,000) |
| | $r(1, C)$ | 0,993 (1,000) | 0,964 (0,969) | 0,973 (1,000) | 0,997 (1,000) |
| | $r(2, C)$ | 0,994 (1,000) | 0,931 (0,936) | 0,989 (1,000) | 0,999 (1,000) |
| TRDF (E4) | $r(1, 2)$ | 0,999 (1,000) | 0,326 (0,338) | 0,838 (1,000) | 0,897 (0,955) |
| | $r(1, C)$ | 1,000 (1,000) | 0,839 (0,852) | 0,958 (1,000) | 0,976 (1,000) |
| | $r(2, C)$ | 1,000 (1,000) | 0,721 (0,737) | 0,955 (1,000) | 0,971 (1,000) |
| CLOR (E5) | $r(1, 2)$ | 0,851 (0,900) | 0,765 (0,772) | 0,955 (1,000) | 0,990 (1,000) |
| | $r(1, C)$ | 0,957 (0,992) | 0,943 (0,948) | 0,986 (1,000) | 0,997 (1,000) |
| | $r(2, C)$ | 0,965 (1,000) | 0,903 (0,907) | 0,987 (1,000) | 0,997 (1,000) |
| ENTC (E6) | $r(1, 2)$ | 0,982 (1,000) | 0,825 (0,837) | 0,914 (1,000) | 0,980 (1,000) |
| | $r(1, C)$ | 0,996 (1,000) | 0,980 (0,985) | 0,981 (1,000) | 0,997 (1,000) |
| | $r(2, C)$ | 0,995 (1,000) | 0,903 (0,913) | 0,967 (1,000) | 0,992 (0,997) |
| APPR (E7) | $r(1, 2)$ | 0,916 (0,965) | 0,550 (0,567) | 0,886 (1,000) | 0,945 (0,980) |
| | $r(1, C)$ | 0,987 (1,000) | 0,894 (0,917) | 0,979 (1,000) | 0,991 (1,000) |
| | $r(2, C)$ | 0,968 (1,000) | 0,826 (0,848) | 0,956 (1,000) | 0,980 (1,000) |
| TRPR (E8) | $r(1, 2)$ | 0,804 (1,000) | 0,453 (0,470) | 0,883 (1,000) | 0,955 (0,989) |
| | $r(1, C)$ | 0,927 (1,000) | 0,863 (0,886) | 0,967 (1,000) | 0,988 (1,000) |
| | $r(2, C)$ | 0,967 (1,000) | 0,818 (0,843) | 0,969 (1,000) | 0,988 (1,000) |
| EVAL (E9) | $r(1, 2)$ | 0,996 (0,996) | 0,305 (0,316) | 0,851 (0,995) | 0,895 (0,942) |
| | $r(1, C)$ | 0,999 (0,999) | 0,658 (0,674) | 0,970 (1,000) | 0,972 (1,000) |
| | $r(2, C)$ | 0,999 (0,999) | 0,816 (0,850) | 0,946 (1,000) | 0,974 (1,000) |

Se destaca una notable invarianza de las medidas de calidad de la enseñanza de los docentes respecto a los ítems para todas las sub-escalas consideradas. Se observa también una alta invarianza de las medidas de los ítems respecto a la faceta estudiantes. La invarianza de las medidas de la calidad de la enseñanza de los docentes respecto a los estudiantes resultó menor que las otras invarianzas consideradas (véase Tabla 6.9). Estos resultados son compatibles también con la hipótesis propuesta, ya que, la menor invarianza observada de los docentes respecto a los estudiantes, se explica a partir de la hipótesis de investigación como una consecuencia de las interacciones de características de

los estudiantes y características de la enseñanza.

6.3.4. Interacciones (sesgos) entre los elementos de las facetas.

El análisis de las interacciones entre las facetas provee de una evaluación más directa de las predicciones teóricas realizadas a partir de la hipótesis de investigación propuesta, en comparación con los resultados presentados en la sección 7.3.2. Aquí se evalúa directamente todas las interacciones posibles entre las facetas docentes, estudiantes e ítems. Adicionalmente, se introduce una faceta “dummy” con medida nula, para investigar las interacción con la enseñanza considerada a nivel de la actividad de enseñanza (curso). La principal predicción teórica a evaluar es que la faceta estudiantes debería tener un menor ajuste al modelo que las facetas docentes e ítems relacionada con cambios en la severidad del estudiante que serían resultado de la interacción entre características del estudiante y las características de la enseñanza. Por tanto, se espera encontrar interacciones de la faceta estudiantes con las facetas docentes y curso.

Tabla 6.10: Análisis de Interacciones (sesgos) entre facetas para cada dimensión considerada. En la Tabla: C (curso), D (docente), E (Estudiante), I (Ítem), * (interacción), Var. Exp (%) (porcentaje de la varianza explicada por la interacción).

| | C*E | | C*I | | D*I | | D*E | | E*I | |
|----------|---------------|--------------------------------|---------------|--------------------------------|---------------|--------------------------------|---------------|--------------------------------|---------------|--------------------------------|
| | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) |
| VALR(E1) | 9,86 | 0,00 | 0,57 | 0,00 | 0,86 | 0,00 | 18,52 | 0,00 | 11,29 | 1 |
| INTR(E2) | 10,37 | 0,00 | 0,76 | 0,00 | 1,22 | 0,00 | 19,13 | 0,00 | 12,29 | 1 |
| RELP(E3) | 11,46 | 0,00 | 0,32 | 0,00 | 0,65 | 0,00 | 20,19 | 0,00 | 7,69 | 1 |
| TRDF(E4) | 10,63 | 0,00 | 1,16 | 0,00 | 1,61 | 0,00 | 17,27 | 0,00 | 13,42 | 1 |
| CLOR(E5) | 9,78 | 0,00 | 0,14 | 0,00 | 0,66 | 0,00 | 17,77 | 0,00 | 9,37 | 1 |
| ENTC(E6) | 10,50 | 0,00 | 0,35 | 0,00 | 1,27 | 0,00 | 19,89 | 0,00 | 10,84 | 1 |
| APPR(E7) | 12,63 | 0,00 | 0,39 | 0,00 | 0,67 | 0,00 | 21,86 | 0,00 | 9,82 | 1 |
| TRPR(E8) | 11,84 | 0,00 | 0,12 | 0,00 | 0,39 | 1 | 11,50 | 0,00 | 16,62 | 0,00 |
| EVAL(E9) | 7,19 | 0,00 | 1,12 | 0,00 | 1,53 | 0,0 | 7,07 | 0,74 | 30,36 | 1 |

En la Tabla 6.10 se resumen los resultados del análisis de interacciones entre las facetas para cada una de las medidas obtenidas a partir de las sub-escalas. Para cada interacción estudiada se presenta una medida del tamaño del efecto

considerado y una medida de significación estadística. Como se aprecia en la Tabla 6.10, considerando las interacciones estadísticamente significativas, el porcentaje de la varianza explicada resultó mayor en general para la interacción Docente*Estudiante, seguida de la interacción Curso*Estudiante. Las demás interacciones, en general no fueron estadísticamente significativas, o bien, el porcentaje de la varianza explicada por la interacción fue pequeño.

Más allá del patrón general encontrado, que resultó consistente con las predicciones teóricas realizadas, se observó una interacción, no predicha, estadísticamente significativa que explica un porcentaje elevado de la varianza observada, para la medida obtenida a partir de la sub-escala TRPR (E8). Para la medida derivada de la sub-escala TRPR (E8) la interacción Estudiante*Item explicó un 16,62% de la varianza y resultó altamente significativa (valor $p < 0,01$). Esta interacción podría ser explicada por la hipótesis de investigación propuesta. Esta interacción podría reflejar cambios en la interpretación de ítems de la escala TRPR (vinculada con tareas de aprendizaje propuestas), relacionadas con diferencias individuales que median la evaluación de las tareas propuestas, o bien, que explican efectos diferentes de las tareas propuestas en el aprendizaje de los estudiantes.

Por otra parte, para la sub-escala EVAL (E9), una de las interacciones predichas (Docente*Estudiante) resultó no estadísticamente significativa y explicó un porcentaje menor de la varianza observada, que para las interacciones observadas para todas las demás sub-escalas. Una posible explicación, a este resultado que parece alejarse del patrón general encontrado, podría ser que en las actividades de enseñanza evaluadas, muchas de las características de la evaluación no son definidas por el docente que realiza el trabajo de aula. Esta explicación sería consistente con la hipótesis de investigación propuesta y con la interacción estadísticamente significativa Curso*Estudiantes que explicó un 7,19% de la varianza.

6.3.5. Estructura de las escalas de los ítems y uso de las categorías.

Las escalas de puntajes de los ítems son la base para la obtención de los datos utilizados. Por tanto, un funcionamiento efectivo de las escalas de puntajes es fundamental para la validez de las inferencias realizadas a partir de las sub-escalas consideradas. La efectividad de las escalas de puntajes de los ítems

está relacionada con la claridad conceptual de los enunciados de los ítems y las categorías definidas. También, está relacionada con la interpretación que realizan los evaluadores de la escala de puntajes.

Si puntajes mayores en la escala de puntajes corresponden a valores mayores de la variable latente se espera que la media de las medidas de las observaciones para cada categoría estén ordenadas en forma monótona creciente. Esto es, las medidas promedio de las observaciones se deberían incrementar desde las categorías de puntajes menores a puntajes mayores. Este patrón aporta evidencias de que a mayor puntaje observado, mayor es el valor de la variable latente. Cómo se muestra en la Tabla 6.11 para todas las sub-escalas se encontró que la media de las medidas de las observaciones presentaron un orden monótono creciente.



Tabla 6.11: Estructura de las escalas de los ítems: Promedio de Medidas

| Items | 1 | 2 | 3 | 4 | 5 |
|-------|-------|-------|-------|------|------|
| VALR1 | -3,37 | -1,17 | 0,61 | 2,94 | 4,93 |
| VALR2 | -3,71 | -1,17 | 0,82 | 3,13 | 5,24 |
| VALR3 | -3,70 | -1,06 | 0,68 | 2,95 | 5,03 |
| VALR4 | -3,53 | -1,29 | 0,74 | 3,11 | 5,07 |
| VALR5 | -3,73 | -0,94 | 0,62 | 2,92 | 4,96 |
| INTR1 | -3,06 | -0,59 | 0,52 | 2,46 | 4,15 |
| INTR2 | -3,12 | -0,93 | 0,71 | 2,72 | 4,50 |
| INTR3 | -3,16 | -0,94 | 0,76 | 2,74 | 4,51 |
| INTR4 | -3,23 | -0,49 | 0,59 | 2,37 | 4,09 |
| INTR5 | -3,13 | -0,82 | 0,79 | 2,79 | 4,59 |
| INTR6 | -2,56 | -0,75 | 0,95 | 2,85 | 4,53 |
| RELP1 | -4,52 | -1,18 | 0,90 | 3,34 | 6,01 |
| RELP2 | -4,58 | -1,64 | 1,26 | 3,78 | 6,58 |
| RELP3 | -4,87 | -1,38 | 0,97 | 3,45 | 6,13 |
| RELP4 | -4,62 | 0,88 | 0,93 | 3,31 | 5,99 |
| RELP5 | -4,81 | -1,35 | 1,02 | 3,54 | 6,24 |
| TRDF1 | -3,24 | -1,10 | 0,81 | 2,13 | 3,71 |
| TRDF2 | -4,05 | -1,54 | 0,38 | 2,07 | 3,70 |
| TRDF3 | -3,80 | -1,38 | 0,35 | 1,88 | 3,58 |
| TRDF4 | -3,67 | -1,38 | 0,47 | 2,01 | 3,47 |
| CLOR1 | -3,79 | -1,05 | 0,68 | 3,17 | 5,60 |
| CLOR2 | -4,06 | -1,09 | 0,71 | 3,17 | 5,66 |
| CLOR3 | -4,20 | -1,20 | 0,91 | 3,26 | 5,73 |
| CLOR4 | -4,10 | -0,97 | 0,93 | 3,38 | 5,89 |
| CLOR5 | -4,14 | -0,92 | 0,95 | 3,39 | 5,91 |
| CLOR6 | -4,17 | -1,13 | 0,85 | 3,31 | 5,81 |
| CLOR7 | -4,18 | -0,93 | 0,96 | 3,33 | 5,82 |
| CLOR8 | -4,14 | -1,12 | 1,13 | 3,56 | 6,04 |
| ENTC1 | -3,22 | -0,68 | 0,86 | 2,83 | 4,86 |
| ENTC2 | -3,34 | -0,63 | 0,86 | 2,78 | 4,82 |
| ENTC3 | -3,47 | -0,32 | 0,76 | 2,52 | 4,47 |
| ENTC4 | -3,26 | -0,20 | 0,83 | 2,66 | 4,73 |
| ENTC5 | -3,15 | -0,62 | 0,90 | 2,79 | 4,75 |
| APPR1 | -3,69 | -0,97 | 0,88 | 3,12 | 5,09 |
| APPR2 | -3,76 | -1,06 | 0,85 | 3,07 | 5,06 |
| APPR3 | -3,88 | -0,86 | 0,82 | 2,99 | 4,93 |
| APPR4 | -3,70 | -1,07 | 0,81 | 3,04 | 5,07 |
| APPR5 | -3,27 | -0,81 | 0,96 | 3,13 | 5,12 |
| TRPR1 | -4,49 | -2,60 | 0,27 | 3,68 | 6,41 |
| TRPR2 | -4,98 | -2,54 | 0,33 | 3,66 | 6,58 |
| TRPR3 | -4,76 | -2,09 | 0,39 | 3,55 | 6,38 |
| TRPR4 | -4,65 | -2,41 | 0,36 | 3,71 | 6,45 |
| EVAL1 | -2,50 | -1,12 | 0,20 | 1,51 | 3,19 |
| EVAL2 | -2,60 | -1,17 | -0,19 | 1,21 | 2,94 |
| EVAL3 | -2,60 | -1,16 | 0,06 | 1,34 | 3,08 |
| EVAL4 | -2,42 | -1,23 | -0,44 | 0,94 | 2,51 |
| EVAL5 | -2,64 | -0,89 | 0,24 | 1,49 | 3,29 |
| EVAL6 | -2,36 | -0,55 | 0,75 | 1,79 | 3,01 |
| EVAL7 | -2,23 | -1,07 | -0,35 | 0,91 | 2,53 |

Si una categoría de la escala de puntajes corresponde a un segmento muy estrecho de la variable latente; o bien, si la variable o concepto tiene una representación mental pobre en los evaluadores, los umbrales pueden estar desordenados (Linacre y cols., 2002). Los umbrales para todos los ítems de las subescalas mostraron un ordenamiento adecuado, con un incremento monótono,

como se aprecia en la Tabla 6.12 y Figura 6.1a. Esto sugiere que las representaciones mentales de los estudiantes y su interpretación de los enunciados de los ítems serían razonablemente adecuadas.

Adicionalmente la distancia entre los umbrales nos informa sobre el funcionamiento de las escalas de puntajes y la interpretación que realizan los evaluadores de la misma (Linacre y cols., 2002). Distancias muy grandes entre dos umbrales consecutivos corresponden con una categoría de la escala de puntajes pobremente definida, en la cual existe una gran dispersión en las medidas de las observaciones en dicha categoría. En este tipo de casos se espera que la categoría considerada aporte poca información en el intervalo central de las medidas. Asimismo, una distancia entre umbrales contiguos no menor que 1 ni mayor que 5 (para escalas de puntajes con cinco categorías) se considera como indicador de calidad de la escala de puntajes.

Un 98 % de los umbrales resultaron mayores que 1 y menores que 5 logits. Solo 3 ítems de la escala EVAL(E7) presentaron una distancia entre los dos primeros umbrales menor que 1 ($u_3 - u_2$), mientras que solo un ítem de la escala RELP(E3) presentó una distancia escasamente mayor a 5 ($u_4 - u_3 = 5,02$).

Tabla 6.12: Estructura de las escalas de los ítems: umbrales de Rasch-Andrich

| Items | 2 | 3 | 4 | 5 |
|-------|--------------|--------------|-------------|-------------|
| VALR1 | -4,14 (0,15) | -2,67 (0,07) | 1,69 (0,03) | 5,12 (0,04) |
| VALR2 | -4,02 (0,19) | -2,51 (0,08) | 1,62 (0,03) | 4,92 (0,04) |
| VALR3 | -4,09 (0,16) | -2,41 (0,07) | 1,70 (0,03) | 4,80 (0,04) |
| VALR4 | -4,19 (0,18) | -2,66 (0,07) | 1,58 (0,03) | 5,27 (0,04) |
| VALR5 | -4,04 (0,15) | -2,58 (0,07) | 1,73 (0,03) | 4,89 (0,04) |
| INTR1 | -3,50 (0,13) | -2,17 (0,06) | 1,48 (0,03) | 4,18 (0,04) |
| INTR2 | -3,64 (0,16) | -2,44 (0,07) | 1,59 (0,03) | 4,48 (0,04) |
| INTR3 | -3,53 (0,17) | -2,43 (0,07) | 1,58 (0,03) | 4,39 (0,04) |
| INTR4 | -3,65 (0,13) | -2,28 (0,06) | 1,66 (0,03) | 4,27 (0,04) |
| INTR5 | -3,97 (0,18) | -2,35 (0,07) | 1,65 (0,03) | 4,67 (0,04) |
| INTR6 | -3,47 (0,19) | -2,44 (0,08) | 1,66 (0,03) | 4,24 (0,04) |
| RELP1 | -4,39 (0,18) | -2,67 (0,08) | 2,00 (0,03) | 5,07 (0,04) |
| RELP2 | -4,86 (0,26) | -2,83 (0,10) | 2,19 (0,04) | 5,51 (0,04) |
| RELP3 | -4,54 (0,20) | -2,54 (0,08) | 2,03 (0,04) | 5,05 (0,04) |
| RELP4 | -4,65 (0,18) | -2,64 (0,07) | 2,16 (0,03) | 5,13 (0,04) |
| RELP5 | -4,64 (0,21) | -2,62 (0,09) | 2,09 (0,04) | 5,17 (0,04) |
| TRDF1 | -4,18 (0,11) | -1,45 (0,04) | 1,99 (0,03) | 3,64 (0,04) |
| TRDF2 | -4,71 (0,10) | -1,70 (0,04) | 2,07 (0,03) | 4,33 (0,06) |
| TRDF3 | -4,50 (0,09) | -1,58 (0,04) | 1,88 (0,03) | 4,19 (0,06) |
| TRDF4 | -4,30 (0,09) | -1,52 (0,04) | 2,03 (0,03) | 3,78 (0,05) |
| CLOr1 | -4,24 (0,17) | -2,39 (0,07) | 1,70 (0,03) | 4,93 (0,04) |
| CLOr2 | -4,33 (0,17) | -2,42 (0,07) | 1,73 (0,03) | 5,03 (0,04) |
| CLOr3 | -4,48 (0,19) | -2,49 (0,07) | 1,99 (0,03) | 4,98 (0,04) |
| CLOr4 | -4,49 (0,21) | -2,64 (0,08) | 1,97 (0,03) | 5,16 (0,04) |
| CLOr5 | -4,68 (0,21) | -2,50 (0,08) | 2,00 (0,03) | 5,18 (0,04) |
| CLOr6 | -4,68 (0,19) | -2,48 (0,07) | 1,97 (0,03) | 5,19 (0,04) |
| CLOr7 | -4,75 (0,20) | -2,64 (0,07) | 2,16 (0,03) | 5,23 (0,04) |
| CLOr8 | -5,24 (0,25) | -2,47 (0,08) | 2,29 (0,03) | 5,43 (0,04) |
| ENTC1 | -3,73 (0,19) | -2,24 (0,07) | 1,71 (0,03) | 4,26 (0,04) |
| ENTC2 | -3,83 (0,18) | -2,12 (0,07) | 1,77 (0,03) | 4,17 (0,04) |
| ENTC3 | -3,82 (0,15) | -1,92 (0,06) | 1,85 (0,03) | 3,89 (0,04) |
| ENTC4 | -4,29 (0,18) | -2,09 (0,06) | 1,93 (0,03) | 4,45 (0,04) |
| ENTC5 | -3,39 (0,18) | -2,34 (0,08) | 1,64 (0,03) | 4,10 (0,03) |
| APPR1 | -4,57 (0,21) | -2,44 (0,07) | 1,87 (0,03) | 5,14 (0,04) |
| APPR2 | -4,66 (0,20) | -2,54 (0,07) | 1,94 (0,03) | 5,27 (0,04) |
| APPR3 | -4,33 (0,18) | -2,52 (0,07) | 1,88 (0,03) | 4,96 (0,04) |
| APPR4 | -4,24 (0,19) | -2,52 (0,07) | 1,81 (0,03) | 4,95 (0,04) |
| APPR5 | -4,33 (0,21) | -2,42 (0,07) | 1,85 (0,03) | 4,89 (0,04) |
| TRPR1 | -5,38 (0,15) | -2,96 (0,06) | 1,72 (0,04) | 6,61 (0,05) |
| TRPR2 | -5,62 (0,15) | -2,92 (0,06) | 1,85 (0,04) | 6,69 (0,05) |
| TRPR3 | -5,46 (0,15) | -2,87 (0,06) | 1,76 (0,04) | 6,58 (0,05) |
| TRPR4 | -5,28 (0,15) | -2,99 (0,06) | 1,87 (0,04) | 6,39 (0,05) |
| EVAL1 | -3,01 (0,09) | -1,82 (0,04) | 1,09 (0,03) | 3,75 (0,05) |
| EVAL2 | -2,94 (0,06) | -1,65 (0,03) | 0,97 (0,03) | 3,61 (0,06) |
| EVAL3 | -3,03 (0,08) | -1,83 (0,04) | 1,02 (0,03) | 3,84 (0,06) |
| EVAL4 | -2,57 (0,04) | -1,54 (0,03) | 1,09 (0,03) | 3,03 (0,07) |
| EVAL5 | -2,89 (0,10) | -2,10 (0,05) | 0,91 (0,03) | 4,07 (0,05) |
| EVAL6 | -2,86 (0,13) | -1,98 (0,06) | 0,94 (0,03) | 3,90 (0,04) |
| EVAL7 | -2,57 (0,05) | -1,70 (0,03) | 0,82 (0,03) | 3,45 (0,06) |

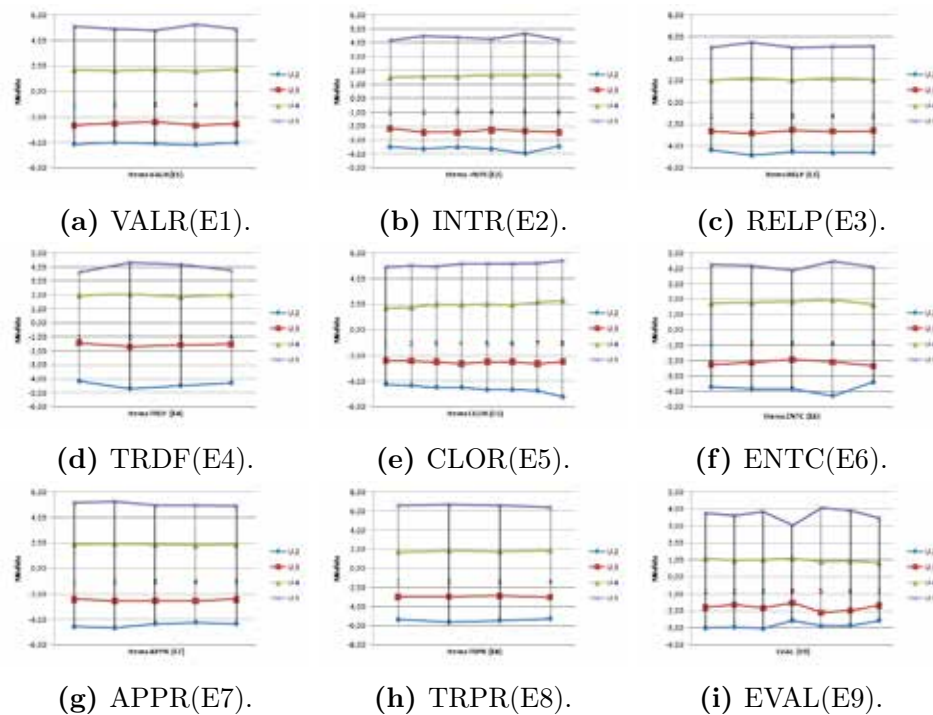


Figura 6.1: Umbrales de Rasch-Aldrich.

El modelo de Rasch permite evaluar el nivel de aleatoriedad observado en los datos. Si la aleatoriedad observada en los datos asociada a cada categoría es la esperada de acuerdo al modelo, se espera que el Outfit sea igual a 1. Valores mayores que uno indican que la aleatoriedad es mayor a la esperada, mientras que, valores menores que uno indican que la aleatoriedad es menor a la esperada. Si el Outfit resulta mayor que 2, la aleatoriedad no explicada es mayor a la explicada, lo que indica que una parte de los datos no es adecuada para obtener mediciones.

En la Tabla 6.13 se muestran los valores de Outfit para las categorías de puntajes de los ítems de las sub-escalas. Todos los valores de Outfit de las categorías resultaron aceptables (menores que 2), lo que significa un ajuste aceptable de las categorías al modelo. No obstante, las categorías de las escalas RELP(E3), CLOR(E5) y ENTC(E6) mostraron un sobre ajuste considerable, esto es, menor aleatoriedad que la predicha por el modelo.

Tabla 6.13: Estructura de las escalas de los ítems: Outfit y Conteo de observaciones no extremas.

| Items | Outfit | | | | | Conteo | | | | |
|-------|--------|-----|-----|-----|-----|--------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| VALR1 | 1,3 | 1,2 | 1,0 | 1,0 | 1,1 | 66 | 311 | 2966 | 3307 | 1016 |
| VALR2 | 0,7 | 1,0 | 0,8 | 0,9 | 1,0 | 42 | 221 | 2353 | 3476 | 1574 |
| VALR3 | 0,8 | 1,1 | 0,9 | 0,8 | 1,0 | 61 | 322 | 2741 | 3210 | 1332 |
| VALR4 | 0,9 | 1,1 | 1,0 | 1,0 | 1,1 | 47 | 245 | 2595 | 3635 | 1144 |
| VALR5 | 0,7 | 1,3 | 0,9 | 0,9 | 1,0 | 68 | 315 | 2909 | 3196 | 1178 |
| INTR1 | 0,9 | 1,4 | 0,9 | 0,9 | 1,0 | 94 | 412 | 2877 | 3041 | 1198 |
| INTR2 | 0,9 | 1,0 | 0,9 | 0,8 | 1,0 | 57 | 272 | 2786 | 3261 | 1246 |
| INTR3 | 0,8 | 1,0 | 0,9 | 0,8 | 1,0 | 55 | 250 | 2661 | 3260 | 1396 |
| INTR4 | 0,8 | 1,5 | 1,2 | 1,0 | 1,1 | 95 | 436 | 3220 | 2845 | 1026 |
| INTR5 | 0,9 | 1,1 | 0,9 | 0,9 | 1,0 | 44 | 284 | 2795 | 3331 | 1168 |
| INTR6 | 1,2 | 1,1 | 0,9 | 0,9 | 1,1 | 43 | 200 | 2467 | 3165 | 1747 |
| RELP1 | 1,1 | 0,0 | 0,0 | 0,0 | 0,0 | 59 | 224 | 2453 | 2928 | 1555 |
| RELP2 | 0,2 | 0,0 | 0,0 | 0,0 | 0,0 | 27 | 134 | 2131 | 3150 | 1777 |
| RELP3 | 0,4 | 0,0 | 0,0 | 0,0 | 0,0 | 48 | 216 | 2307 | 2917 | 1731 |
| RELP4 | 1,0 | 0,0 | 0,0 | 0,0 | 1,1 | 58 | 263 | 2711 | 2793 | 1394 |
| RELP5 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 42 | 195 | 2300 | 2968 | 1714 |
| TRDF1 | 1,2 | 1,0 | 0,9 | 1,1 | 1,1 | 132 | 919 | 3568 | 2148 | 1046 |
| TRDF2 | 0,9 | 0,9 | 0,8 | 0,9 | 1,0 | 167 | 1262 | 4252 | 1747 | 385 |
| TRDF3 | 1,1 | 1,1 | 0,9 | 1,0 | 1,0 | 198 | 1328 | 3999 | 1860 | 428 |
| TRDF4 | 1,0 | 1,0 | 0,9 | 0,9 | 1,1 | 203 | 1263 | 4004 | 1742 | 601 |
| CLOR1 | 0,2 | 0,0 | 0,0 | 0,0 | 0,0 | 59 | 296 | 2352 | 3129 | 1520 |
| CLOR2 | 0,1 | 0,0 | 0,0 | 0,0 | 0,0 | 56 | 295 | 2393 | 3166 | 1446 |
| CLOR3 | 0,2 | 0,0 | 0,0 | 0,0 | 0,0 | 45 | 266 | 2524 | 2971 | 1550 |
| CLOR4 | 0,3 | 0,0 | 0,0 | 0,0 | 0,0 | 37 | 213 | 2412 | 3127 | 1567 |
| CLOR5 | 0,8 | 0,0 | 0,0 | 0,0 | 0,0 | 34 | 239 | 2425 | 3110 | 1548 |
| CLOR6 | 0,2 | 0,0 | 0,0 | 0,0 | 0,0 | 42 | 282 | 2576 | 3086 | 1370 |
| CLOR7 | 0,1 | 0,0 | 0,0 | 0,0 | 0,0 | 39 | 259 | 2770 | 2952 | 1336 |
| CLOR8 | 0,2 | 0,0 | 0,0 | 0,0 | 0,0 | 23 | 251 | 2671 | 3029 | 1382 |
| ENTC1 | 1,0 | 0,0 | 0,0 | 0,0 | 0,9 | 44 | 244 | 2492 | 3000 | 1680 |
| ENTC2 | 0,9 | 0,0 | 0,0 | 0,0 | 0,9 | 46 | 288 | 2590 | 2873 | 1663 |
| ENTC3 | 0,8 | 0,0 | 0,0 | 0,0 | 1,0 | 70 | 445 | 2900 | 2523 | 1522 |
| ENTC4 | 1,2 | 0,0 | 0,0 | 0,0 | 1,0 | 48 | 400 | 3082 | 2758 | 1172 |
| ENTC5 | 0,9 | 0,0 | 0,0 | 0,0 | 1,0 | 49 | 207 | 2352 | 2970 | 1882 |
| APPR1 | 0,7 | 1,1 | 0,9 | 0,9 | 1,0 | 32 | 291 | 2933 | 3457 | 1189 |
| APPR2 | 0,6 | 1,1 | 1,0 | 1,0 | 1,0 | 35 | 312 | 3197 | 3367 | 991 |
| APPR3 | 0,6 | 1,2 | 1,0 | 1,0 | 1,0 | 44 | 307 | 3096 | 3274 | 1181 |
| APPR4 | 0,6 | 1,0 | 0,9 | 0,9 | 1,0 | 41 | 274 | 2903 | 3397 | 1287 |
| APPR5 | 1,0 | 1,1 | 0,9 | 0,9 | 1,0 | 31 | 251 | 2718 | 3396 | 1506 |
| TRPR1 | 1,1 | 0,8 | 0,8 | 0,9 | 1,1 | 68 | 436 | 2891 | 3636 | 764 |
| TRPR2 | 0,9 | 0,8 | 0,9 | 0,9 | 0,9 | 63 | 478 | 3037 | 3521 | 696 |
| TRPR3 | 0,9 | 1,1 | 1,1 | 1,1 | 1,1 | 67 | 469 | 2913 | 3581 | 765 |
| TRPR4 | 0,9 | 0,8 | 0,9 | 0,8 | 1,0 | 66 | 403 | 2956 | 3472 | 898 |
| EVAL1 | 0,8 | 0,8 | 0,8 | 0,9 | 1,0 | 184 | 798 | 3533 | 2607 | 578 |
| EVAL2 | 0,8 | 1,0 | 0,8 | 0,9 | 0,9 | 393 | 1304 | 3523 | 2094 | 386 |
| EVAL3 | 0,8 | 0,8 | 0,8 | 1,0 | 1,0 | 240 | 947 | 3609 | 2459 | 445 |
| EVAL4 | 1,0 | 1,2 | 1,0 | 0,9 | 0,9 | 821 | 1705 | 3375 | 1440 | 359 |
| EVAL5 | 0,7 | 0,9 | 0,9 | 1,0 | 1,0 | 153 | 570 | 3358 | 3081 | 538 |
| EVAL6 | 0,8 | 1,0 | 1,1 | 1,2 | 1,4 | 82 | 377 | 2797 | 3509 | 935 |
| EVAL7 | 1,2 | 1,4 | 1,1 | 1,0 | 1,1 | 605 | 1341 | 3375 | 2010 | 369 |

En la Tabla 6.13 se muestran también las frecuencias de entre las categorías. Las distribuciones de las frecuencias de respuestas de las categorías encontradas resultaron regulares, sin la presencia de categorías no observadas.

Esto representa una evidencia adicional de la calidad de las escalas de puntajes (Bond y Fox, 2015; Eckes, 2009; Linacre y cols., 2002).



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

6.4. Discusión.

La presente investigación sigue una orientación al estudio de la validez de los SET que es similar a la fundada en el análisis de la representación de constructo (Embretson, 1994) y el enfoque substantivo a la validez de constructo (Messick, 1995). El enfoque considerado en esta investigación en relación a la validez de constructo se basa en el estudio teórico y empírico del proceso de respuesta desde una perspectiva cognitiva y socio-cognitiva.

La hipótesis central que se propuso evaluar aquí es que existen diferencias individuales en cómo los estudiantes evalúan la enseñanza que están relacionadas con variables que median la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza, o bien, que explican diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes.

Específicamente, varias de las variables del estudiante que pueden mediar la interpretación cognitiva, o bien, explicar diferencias en cómo la enseñanza facilita el aprendizaje, son variables de estado, más que atributos. Estas, dependen del estudiante, de la enseñanza impartida por el docente y el contexto. Por el contrario la efectividad o calidad de la enseñanza del docente, así como también, la dificultad de satisfacer los criterios de evaluación, se espera que sean atributos razonablemente estables. Por ejemplo, la motivación intrínseca de un estudiante para realizar una tarea de aprendizaje depende de características de la tarea y del estudiante.

Por tanto, si el estudiante está expuesto a enseñanzas con diferentes características, se espera que variables como la motivación intrínseca, la autoeficacia para aprender o las dimensiones de causalidad atribuida a éxitos y fracasos en un curso, cambien de una actividad de enseñanza a otra con características diferentes. Por tanto, una consecuencia que se deduce de la observación anterior y de la hipótesis central de esta investigación, es que se esperan cambios en el proceso de evaluación de los estudiantes, en particular, de su severidad al evaluar la enseñanza. Esta predicción teórica es consistente con diversas investigaciones previas (R. Ames y Lau, 1979; Bandura y cols., 1996; Grimes y cols., 2004; Svanum y Aigner, 2011; Wigfield y cols., 1997; Zimmerman, 2000).

Las predicciones teóricas realizadas pueden expresarse en términos de modelos de múltiples facetas de Rasch. En términos de este modelo se espera que existan cambios en la severidad de los estudiantes que serían resultado de la interacción entre características del estudiante y características de la enseñan-

za. A su vez, esto implica que el ajuste esperado entre los datos y el modelo MFRM solo con efectos primarios (sin interacciones) sea menor en la faceta estudiantes que en las facetas ítems y docentes. Además, se espera también que la invarianza de las medidas de los docentes respecto a los estudiantes, para cada sub-escala, sean menores que las demás invarianzas consideradas. Estas predicciones teóricas fueron ampliamente confirmadas por los resultados obtenidos.

El patrón general encontrado fue que considerando las interacciones estadísticamente significativas, el porcentaje de la varianza explicada resultó en general mayor para la interacción Docente*Estudiante, seguida de la interacción Curso*Estudiante. Adicionalmente, en general las demás interacciones no fueron estadísticamente significativas, o bien, el porcentaje de la varianza explicada por la interacción fue pequeño. Este resultado, es especialmente destacable, debido a que este patrón general se verificó para dimensiones relacionadas, pero, distintas de calidad en la enseñanza. En relación a las invarianzas de las medidas, los resultados confirmaron claramente la predicción realizada. Las invarianzas obtenidas para las medidas de los docentes respecto a los estudiantes fueron, para todas las sub-escalas, menores que las demás invarianzas.

Una excepción al patrón general encontrado estuvo relacionada con la sub-escala TRPR(E8). Esta escala está vinculada con las tareas de aprendizaje propuestas por el docente. Para esta escala también se encontraron interacciones Docente*Estudiante y Curso*Estudiante estadísticamente significativas que explicaron un 11,50 % y un 11,84 % de la varianza respectivamente, consistente con las predicciones teóricas realizadas. Sin embargo, la escala TRPR(E8) presentó una interacción estadísticamente significativa entre las facetas Estudiantes*Ítems que explicó un 16,62 % de la varianza. La interacción Estudiantes*Ítems muestra que en este caso no solo se produjo un cambio en la severidad de los estudiantes al evaluar a los docentes, sino que también, un cambio en la forma de interpretación de los ítems de la escala. Esto refleja un cambio en el proceso de respuesta, originalmente no predicho, pero que puede ser consistente con la hipótesis central de esta investigación.

La sub-escala TRPR(E8) está constituida por 4 ítems. Los ítems TRPR01 y TRPR02 hacen referencia directa y explícitamente a contenidos vinculados estrechamente con la satisfacción de metas de aprendizaje. Por el contrario, el ítem TRPR03 refiere a la coherencia de las tareas propuestas con otros aspec-

tos del curso; mientras que, el ítem TRPR04 plantea un criterio de evaluación más general, vinculado al valor atribuido a las tareas propuestas. Tomando en cuenta estas observaciones sobre el contenido de los ítems de esta escala, se puede apreciar que la interacción encontrada Estudiantes*Items puede explicarse con la hipótesis central de esta investigación.

Por ejemplo, estudiantes con preferencias por metas de aprendizaje respecto de metas de desempeño, de acuerdo a la hipótesis central propuesta y al modelo de Auto-teorías (véase capítulo 4), se espera que sean menos severos al evaluar a los docentes cuando consideran los criterios definidos por los ítems TRPR01 y TRPR02, respecto de estudiantes que prefieren las metas de desempeño respecto de las metas de aprendizaje. Esta explicación teórica es consistente con el patrón de términos de interacción encontrados entre los estudiantes e ítems, que corresponden a severidades de los estudiantes con correlaciones lineales más fuertes para el par TRPR01 y TRPR02, por un lado, y para el par TRPR03 y TRPR04, por otro (resultado no presentado aquí).

Un aspecto singular de esta investigación, además del abordaje al estudio de la validez previamente mencionado, es el enfoque metodológico empleado. Tradicionalmente, las investigaciones sobre SET consideran como unidad de análisis los promedios de puntajes para cada ítem por grupo de estudiantes que evalúan un docente. Los métodos de análisis más frecuentemente aplicados son análisis correlacionales, algunos modelos de regresión, métodos de análisis factorial exploratorio y confirmatorio (Marsh, 1987; Theall y cols., 2001). Más recientemente, algunas investigaciones han considerado también modelos de ecuaciones estructurales (Marsh y cols., 2009; Paswan y Young, 2002; Shevlin, Banyard, Davies, y Griffiths, 2000).

La presente investigación propone una aproximación metodológica relativamente innovadora fundada en la aplicación de modelos MFRM y un método adecuado de diagnóstico del modelo para estudios observacionales, que permite sortear dificultades inherentes a problemas de conexión en las redes de evaluación implicadas (Engelhard Jr, 1996; Linacre, 2015). Adicionalmente, el diseño adoptado en esta investigación permite estudiar el proceso de respuesta a los SET en condiciones realistas, ecológicamente válidas, para una muestra de estudiantes de considerable tamaño y diversidad, de forma tal que el proceso de respuesta de cada estudiante se observó en forma repetida en diferentes actividades de enseñanza durante dos semestres consecutivos. Este aspecto se considera clave para poder evaluar la hipótesis central de esta investigación.

La perspectiva teórica y metodológica que se adoptó este estudio permitió además caracterizar las propiedades psicométricas del instrumento considerado y aportar varias evidencias de validez que se relacionan con el análisis del proceso de respuesta de los estudiantes (evaluadores).

En primer lugar, los resultados encontrados son consistentes con una adecuada interpretación de los ítems. Los ítems de las sub-escalas en general muestran un elevado ajuste al modelo MFRM sin interacciones, mostrado por los valores de los estadísticos de ajuste, con la excepción de la escala TRPR(E8), que muestra un ajuste menor, pero satisfactorio. Esto en parte estaría relacionado con la interacción Estudiantes*Ítems encontrada para esta sub-escala.

Adicionalmente, los valores encontrados de r_{pm} muestran un nivel alto de consistencia entre las medidas de los ítems y el constructo medido por cada sub-escala. Además, los niveles de confiabilidad de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) muestra que los estudiantes son capaces de apreciar las diferencias que existen de contenido entre los ítems. Esto implica que es poco probable que exista un efecto halo a nivel de grupo (Eckes, 2011; Engelhard Jr, 2013; Myford y Wolfe, 2004).

No obstante, el rango de variación de las dificultades, es relativamente estrecho, y es mucho menor que el rango de variación de las medidas de los docentes y severidades de los estudiantes. Esto representa una limitación para establecer procedimientos de evaluación de la calidad de enseñanza referida al criterio (Bond y Fox, 2015; B. D. Wright y Stone, 1999). También, limita en alguna medida la comprensión del significado de los constructos medidos.

En segundo lugar, los resultados obtenidos muestran que los estudiantes, como grupo, son capaces de distinguir entre los docentes según la magnitud de la calidad o efectividad en la enseñanza en las dimensiones medidas por las sub-escalas estudiadas. Los docentes muestran un buen nivel de ajuste al modelo para todas las sub-escalas consideradas. Adicionalmente, la medida de cada docente es altamente consistente con el constructo medido por la escala, lo que se aprecia al considerar los valores obtenidos de r_{pm} . Además, los niveles de confiabilidad de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) muestra que los estudiantes son capaces de apreciar las diferencias que existen de calidad o efectividad de la enseñanza entre los docentes. Esto implica que es poco probable que exista un efecto de restricción de rango a nivel de grupo (Engelhard Jr, 2013; Myford y Wolfe, 2004).

Por último, los resultados obtenidos en esta investigación muestran que

existen entre 4 y 5 estratos estadísticamente discernibles de severidad de los estudiantes, lo que muestra que existen diferencias importantes en cómo los estudiantes evalúan a los docentes. Esto se ve reflejado también en el amplio rango de variación de las severidades. Si bien los estudiantes presentan diferencias importantes en su severidad al evaluar, varios resultados son consistentes con una aceptable capacidad para evaluar la enseñanza.

Como se mencionó previamente en esta discusión, los estudiantes como grupo son capaces de interpretar adecuadamente los ítems de las escalas. Adicionalmente, los resultados relativos a la estructura y uso de las escalas mostraron que son también capaces de interpretar en forma consistente las escalas de puntajes de los ítems. También, como se explicó previamente en esta discusión, los estudiantes globalmente son capaces de apreciar las diferencias que existen en la calidad o efectividad de los docentes.

Por tanto, a un nivel cognitivo, las evidencias obtenidas parecen apoyar la idea que en términos generales las representaciones mentales de las correlaciones entre señales utilizadas para evaluar la enseñanza y los atributos de calidad en la enseñanza son razonablemente adecuadas (Albright y Malloy, 2001; Engelhard Jr, 2013; Hammond y Stewart, 2001; Tapp, 1984). En forma similar, parece ser que las teorías implícitas acerca de las dimensiones de la enseñanza y cómo estas se relacionan entre sí tienen un nivel de exactitud al menos aceptable (Marsh, 1987; Renaud y Murray, 2005). No obstante, los resultados de ajuste de cada estudiante al modelo indicaron que alrededor de un 10% de los estudiantes mostraron un patrón de respuesta inconsistente y sugieren cierta variación entre los estudiantes en la consistencia y exactitud de sus evaluaciones.

En conclusión, los resultados encontrados aportan evidencias de la validez del cuestionario considerado y son consistentes con las predicciones realizadas. Los resultados obtenidos en esta investigación pueden contribuir a un mejor uso e interpretación de estos instrumentos en la práctica de evaluación de la enseñanza con fines diagnósticos y de mejora de la calidad en la enseñanza.



Capítulo 7

Validez de un cuestionario de evaluación socio-cognitiva del proceso de aprendizaje: análisis de múltiples facetas de Rasch.

7.1. Presentación.

La investigación presentada en este capítulo comprende dos objetivos principales. Primero, busca aportar evidencias de validez de un cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. En segundo lugar, la presente investigación se propone evaluar algunas predicciones teóricas derivadas de la hipótesis central de esta tesis y de resultados reportados en el capítulo previo.

En este capítulo se plantea que el menor ajuste de los estudiantes en relación al modelo MFRM considerado en el capítulo 6, en el que se mostró que estaban implicadas interacciones Docente*Estudiante, podría ser una consecuencia de variaciones en la severidad del estudiante relacionadas con diferentes efectos facilitadores de los docentes.

De acuerdo a los resultados reportados en el capítulo 6: existe una gran variabilidad en la severidad global de los estudiantes al evaluar la calidad de la enseñanza; en segundo lugar, para todas las dimensiones consideradas, existen diferencias en la severidad local de los estudiantes, que se manifestaron como interacciones Docentes*Estudiante.

A partir de estos resultados y la hipótesis principal de esta tesis, en el presente trabajo se deducen dos predicciones teóricas: en primer lugar, en la muestra de calibración estudiada debería existir una considerable variabilidad en los constructos medidos entre los estudiantes; en segundo lugar, debería existir variabilidad en el efecto facilitador de los docentes en relación a cada uno de los constructos considerados. Estas predicciones teóricas fueron evaluadas empíricamente y se discuten las consecuencias teóricas y prácticas de los resultados obtenidos.

7.2. Metodología.

La metodología básica empleada en la presente investigación fue descrita en el capítulo 3 y en el capítulo 5 sección 5.2. En esta sección se describen aspectos específicos correspondientes a este capítulo.

7.2.1. Muestra de datos de calibración.

La muestra de datos de calibración consistió en las respuestas de 3243 estudiantes a 7 sub-escalas del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. Las respuestas se obtuvieron durante dos semestres consecutivos, en 10 actividades de enseñanza, en las que participaron 98 docentes, en la carrera de médico durante el año 2015. Estos estudiantes respondieron también en este período a cuestionarios de evaluación de la calidad de la enseñanza.

7.2.2. Instrumentos.

El instrumento aplicado fue un cuestionario de evaluación socio-cognitiva del proceso de aprendizaje que comprende 7 sub-escalas y un total de 35 ítems (Tabla 7.1). Los principales constructos considerados en este instrumento son: tipos de atribuciones, auto-eficacia para aprender, expectativa de resultado, motivación intrínseca, percepción de elección, valor/utilidad y elección de metas. Estas variables son centrales en teorías cognitivas de la motivación, y están involucradas en el proceso de aprendizaje. El proceso de desarrollo de estos instrumentos fue descrito en el capítulo 3. Una descripción de los constructos considerados, y de las teorías relacionadas con estos, fue expuesta en

el capítulo 2. El papel de estos constructos en relación con la hipótesis central de esta tesis fue descrito en los capítulos 1, 2 y 4.

Tabla 7.1: Descripción del cuestionario socio-cognitivo de evaluación del proceso de aprendizaje

| Sub-Escala | Constructo | Nº ítems |
|------------|------------------------------------------------|----------|
| AUTE (S1) | Auto-eficacia para aprender | 9 |
| ERES (S2) | Expectativa positiva de resultados | 4 |
| MOTI (S3) | Motivación Intrínseca | 4 |
| PREL (S4) | Percepción de Elección | 4 |
| VALU (S5) | Valor/Utilidad | 4 |
| ELEM (S6) | Elección de Métras (aprendizaje vs. desempeño) | 5 |
| ATRI | Atribuciones causales específicas | 5 |
| LOCU (S7) | Locus de causalidad | - |
| ESTA (S8) | Estabilidad | - |
| CONT (S9) | Controlabilidad | - |

7.2.3. Aproximación Analítica.

Las respuestas de los estudiantes a cada sub-escala fue analizada por separado aplicando un modelo de múltiples facetas de Rasch (véase sección 5.2). Para entender la base racional del modelo considerado aquí, se tomará como ejemplo la motivación intrínseca del estudiante. El modelo considerado asume que la motivación intrínseca experimentada por un estudiante (MI_{Exp}) en un curso impartido por cierto docente, resulta de la adición de la motivación intrínseca característica del estudiante (un valor estable del estudiante, o MI_{Est}) más el efecto facilitador del docente sobre la motivación intrínseca (MI_{Doc}). El modelo de efectos primarios asume que la motivación intrínseca en el estudiante resulta de la adición de ambos términos, esto es: $MI_{Exp} = MI_{Est} + MI_{Doc}$.

Cuando además se considera que puede existir una interacción entre la motivación intrínseca estable del estudiante (MI_{Est}) y el efecto facilitador del docente sobre la motivación intrínseca (MI_{Doc}), al modelo anterior se le adicio-

na un término de interacción, esto es: $MI_{Exp} = MI_{Est} + MI_{Doc} + MI_{Est} * MI_{Doc}$. En el último caso, el efecto facilitador del docente sobre la motivación intrínseca experimentada por el estudiante se representa con un efecto aditivo y un efecto multiplicativo. La discusión anterior de carácter conceptual e ilustrativo puede expresarse en términos de modelos de múltiples facetas de Rasch.

El modelo de múltiples facetas aplicado aquí contiene tres facetas de medición (docentes, estudiantes e ítems). La medida del docente se interpreta como el efecto facilitador del docente sobre el constructo estudiado (por ejemplo, efecto facilitador sobre la motivación intrínseca o MI_{Doc}). La medida del estudiante corresponde a un nivel estable del constructo estudiando. Por ejemplo, si se considera la motivación intrínseca, correspondería a un valor característico y estable del estudiante (MI_{Est}). Por último, la medida de los ítems se interpreta como una dificultad, que expresa cierta magnitud del constructo medido, característica del ítem, que los estudiantes deben alcanzar para adherir o acordar con la afirmación expresada en el ítem a cierto nivel.

Entonces, el modelo de medición de múltiples facetas de Rasch (modelo MFRM) utilizado para analizar los datos de respuestas a escalas de evaluación socio-cognitiva del proceso de aprendizaje puede ser especificado como sigue:

$$\ln \left[\frac{P_{nij k}}{P_{nij k-1}} \right] = \theta_n + \alpha_j - \beta_i - \tau_{ik} \quad (7.1)$$

donde

$P_{nij k}$ = Probabilidad de que el estudiante j seleccione un puntaje k en el ítem i en el contexto de una actividad de enseñanza impartida por el docente n ,

$P_{nij k-1}$ = Probabilidad de que el estudiante j seleccione un puntaje $k-1$ en el ítem i en el contexto de una actividad de enseñanza impartida por el docente n ,

θ_n = Efecto facilitador del docente n (por ejemplo, efecto facilitador del docente sobre la motivación intrínseca),

α_j = Medida del constructo (por ejemplo, motivación intrínseca estable) para el estudiante j ,

β_i = Dificultad del ítem i ,

τ_{ik} = Dificultad de recibir un puntaje k relativo a recibir un puntaje $k-1$ para el ítem i .

Los estadísticos básicos empleados en este capítulo fueron detalladamente expuestos en el capítulo 3. En cada análisis se realizó un diagnóstico de los modelos con un enfoque general similar al expuesto en el capítulos 5 y 6. La estructura de puntajes de las escalas y la validez de las categorizaciones fueron cuidadosamente analizadas. En los casos que se encontraron problemas en la calidad de las categorizaciones se aplicaron los criterios discutidos en el capítulo 3 y 6. Cuando fue necesario se aplicaron métodos correctivos como por ejemplo colapsar categorías de puntajes contiguas.

Para obtener una medida de las dimensiones de causalidad se realizaron tres análisis independientes de la sub-escala ATRI. Cada análisis requiere recodificación de algunos de los items de la escala ATRI de forma tal de reflejar categorizaciones teóricas de causas específicas según sus dimensiones de causalidad (Weiner, 1983). Así, a partir de la escala ATRI se redefinieron tres sub-escalas: LOCU (S7), ESTA (S8) y CONT (S9). La metodología empleada puede considerarse una extensión de métodos indirectos utilizados para estimar las dimensiones de causalidad basada en modelos de múltiples facetas de Rasch. Una ventaja importante de esta aproximación, es que ofrece una manera conveniente de evaluar la adecuación de los modelos teóricos sobre las dimensiones causales de causas específicas. También, esta aproximación permite evaluar detalladamente la validez de las medidas obtenidas.

7.3. Resultados.

7.3.1. Ajuste global entre datos y el modelo.

Los resultados reportados en esta sección corresponden al ajuste final global obtenido entre los datos y el modelo con efectos primarios (sin interacciones). Cada sub-escala fue analizada por separado. Durante el proceso de diagnóstico del modelo no fue necesario eliminar elementos de las facetas consideradas. No obstante, fue necesario en algunos casos colapsar categorías de puntajes contiguas.

Se encontró un buen ajuste global entre los datos y el modelo para todas las sub-escalas consideradas (Tabla 7.2). Los valores de *Infit* y *Outfit* para todas las escalas muestran un ajuste satisfactorio (Linacre, 2014). Además, los valores obtenidos de *Infit* y *Outfit* muestran también que la aleatoriedad observada en los patrones de respuesta son consistentes con el modelo.

Tabla 7.2: Ajuste Global entre los datos y el modelo.

| | AUTE | ERES | MOTI | PREL | VALU | ELEM | LOCU | ESTA | CONT |
|----------------------------|--------|--------|--------|--------|--------|--------|-------|--------|--------|
| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 |
| <i>Infit</i> | 1,00 | 1,00 | 0,98 | 0,98 | 0,99 | 0,99 | 1,00 | 0,99 | 0,99 |
| <i>Outfit</i> | 1,04 | 1,03 | 1,13 | 1,05 | 1,19 | 1,07 | 1,01 | 1,05 | 1,03 |
| <i>Media</i> (Z_{Res}) | 0,00 | 0,00 | -0,00 | 0,02 | 0,00 | 0,00 | 0,04 | 0,00 | 0,00 |
| <i>SD</i> (Z_{Res}) | 0,99 | 0,99 | 1,01 | 0,99 | 0,99 | 0,99 | 0,96 | 1,01 | 0,99 |
| PRE_3 | 0,01 | 0,01 | 0,01 | 0,00 | 0,01 | 0,01 | 0,00 | 0,00 | 0,00 |
| PRE_2 | 0,05 | 0,06 | 0,05 | 0,04 | 0,06 | 0,04 | 0,04 | 0,03 | 0,04 |
| PVE | 0,56 | 0,54 | 0,67 | 0,46 | 0,62 | 0,41 | 0,37 | 0,40 | 0,40 |
| MAE | 0,47 | 0,40 | 0,42 | 0,59 | 0,33 | 0,39 | 0,38 | 0,47 | 0,44 |
| RMSE | 0,64 | 0,55 | 0,58 | 0,76 | 0,49 | 0,52 | 0,51 | 0,50 | 0,57 |
| NMAE | 13,04 | 14,39 | 11,86 | 22,19 | 11,63 | 17,42 | 16,20 | 21,70 | 19,71 |
| NMBE | 2,41 | 0,04 | 0,03 | 0,05 | 0,07 | 0,10 | 0,18 | 0,03 | 0,06 |
| N° Est. | 3243 | 3016 | 3033 | 3034 | 3012 | 3015 | 3016 | 3016 | 3016 |
| N° Obs. | 80.082 | 33.440 | 33.596 | 33.596 | 33.396 | 41.795 | 41800 | 41.800 | 33.440 |

Adicionalmente, los valores obtenidos de la media ($Media(Z_{Res})$) y desvío estándar ($SD(Z_{Res})$), de los residuos estandarizados (Z_{Res}) para el modelo, son muy próximos a los valores esperados para un ajuste perfecto entre los datos y el modelo ($Media(Z_{Res}) = 0$ y $SD(Z_{Res}) = 1$). En forma similar las proporciones de residuos estandarizados absolutos extremos (PRE_3 y PRE_2) son muy próximos a los valores esperados para un ajuste perfecto ($PRE_3 = 0,00$ y $PRE_2 = 0,05$).

Por su parte, cuando se analiza el ajuste absoluto global entre los datos y el modelo desde una perspectiva que prioriza la exactitud y precisión en la predicción de los puntajes observados, se encuentra en general un ajuste aceptable. Los valores de MAE mostrados en la Tabla 7.2, muestran que el error absoluto promedio, o diferencia absoluta promedio entre el puntaje observado y el predicho por el modelo, es relativamente bajo. Los valores de MAE resultaron inferiores a un 20% de la media de puntajes observados, excepto para las escalas $PREL$ y $ESTA$ que son algo superiores a dicho valor. Además, los valores de MAE son en general menores a la mitad de la diferencia mínima entre dos puntajes observados diferentes, con la única excepción de la escala $PREL$. Por otra parte, los valores de $NMBE$ (o sesgo promedio de los errores normalizado) muestran que el sesgo promedio de los valores predichos por el modelo son en general muy bajos.

Los valores de PVE en la Tabla 7.2 describen la proporción de la varianza de los puntajes observados explicada por el modelo. Como se aprecia en la Tabla 7.2 los valores de PVE oscilan entorno a un 50% de la varianza de los puntajes observados. Por tanto, una proporción substancial de la varianza observada es explicada por el modelo.

7.3.2. Análisis de las facetas de Docentes, Estudiantes e Ítems.

Faceta Ítems.

En la Tabla 7.3 se muestran las medidas obtenidas para cada ítem, el error estándar de medición y estadísticos de ajuste. Los valores de Infit y Outfit obtenidos para cada ítem indican un alto nivel de ajuste al modelo. Los valores de correlación punto medida (r_{pm}) mostraron consistencia entre las medidas de cada ítem y el constructo medido por su escala, para todos los ítems y sub-escalas consideradas. No obstante, los valores de coeficientes de discriminación (*Discrim*) presentaron en muchos casos valores relativamente alejados del valor 1,00 esperado si el ajuste entre los datos y el modelo es perfecto.

Los valores presentados en la Tabla 7.3 son resumidos para cada sub-escala en la Tabla 7.4, para mostrar la distribución de las medidas y estadísticos de ajuste de los ítems para cada sub-escala. En todos los casos los valores de Infit y Outfit medios son muy cercanos al valor 1 (ajuste perfecto) con un desvío estándar relativamente pequeño. En forma similar el valor de la media de r_{pm} indica un alto nivel de consistencia entre las medidas de los ítems y el constructo medido por la escala, también con un valor del desvío estándar de r_{pm} muy pequeño.

El análisis de la confiabilidad de las medidas de los ítems se presenta en la Tabla 7.5. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 ítems tienen diferente dificultad, para cada una de las sub-escalas consideradas.

En la Tabla 7.5 se muestra los valores de la razón de separación (G) para cada escala. G describe la amplitud de variación de las medidas de dificultad (SD_A) tomando como unidad el error estándar de medición ($RMSE$). Los valores de G muestran que en general las medidas de dificultad de los ítems de las sub-escalas son notablemente confiables.

Esta alta confiabilidad en las mediciones se refleja en el número de estratos estadísticamente discernibles (H) que podrían definirse para las dificultades de los ítems. Por ejemplo, si bien la escala AUTE(S1) contiene nueve ítems, con el nivel de confiabilidad de las mediciones encontrado, se podrían definir cerca de 42 estratos de dificultad. Se encontró también un alto valor de confiabilidad de separación (SR), consistente con los valores obtenidos de los otros índices

de confiabilidad.

Si bien en general la confiabilidad de las mediciones resultó elevada, existen algunas variaciones en la confiabilidad entre escalas. En particular, las medidas de los ítems obtenidas con la sub-escala VALU(S5) presentaron una mucho menor confiabilidad que el resto, mientras que la sub-escala ESTA(S8) presentó el valor más elevado de la confiabilidad.

Tabla 7.3: Ítems : Medidas y estadísticos de ajuste por ítem

| Ítems | Medida | S.E. | Infit | Outfit | r_{pm} | Discrim |
|-------|--------|------|-------|--------|----------|---------|
| AUTE1 | 0,79 | 0,02 | 1,20 | 1,27 | 0,71 | 0,77 |
| AUTE2 | -0,88 | 0,02 | 0,97 | 0,99 | 0,69 | 1,04 |
| AUTE3 | 0,63 | 0,02 | 1,08 | 1,11 | 0,73 | 0,91 |
| AUTE4 | 0,34 | 0,02 | 0,97 | 1,03 | 0,75 | 1,02 |
| AUTE5 | -0,16 | 0,02 | 0,86 | 0,89 | 0,75 | 1,14 |
| AUTE6 | 0,05 | 0,02 | 1,08 | 1,18 | 0,71 | 0,92 |
| AUTE7 | -0,60 | 0,02 | 0,95 | 0,95 | 0,71 | 1,07 |
| AUTE8 | -0,26 | 0,02 | 0,96 | 0,97 | 0,72 | 1,04 |
| AUTE9 | 0,09 | 0,02 | 0,90 | 0,94 | 0,75 | 1,10 |
| ERES1 | 0,32 | 0,02 | 1,31 | 1,42 | 0,67 | 0,62 |
| ERES2 | 0,22 | 0,02 | 0,93 | 0,93 | 0,76 | 1,07 |
| ERES3 | -0,2 | 0,02 | 0,87 | 0,89 | 0,77 | 1,14 |
| ERES4 | -0,34 | 0,02 | 0,86 | 0,87 | 0,78 | 1,16 |
| MOTI1 | -0,24 | 0,02 | 0,9 | 1,05 | 0,82 | 1,08 |
| MOTI2 | 0,22 | 0,02 | 0,99 | 1,15 | 0,81 | 0,99 |
| MOTI3 | 0,5 | 0,02 | 1,12 | 1,28 | 0,8 | 0,86 |
| MOTI4 | -0,48 | 0,02 | 0,91 | 1,02 | 0,81 | 1,08 |
| PREL1 | -0,2 | 0,01 | 0,82 | 0,94 | 0,68 | 1,26 |
| PREL2 | 0,31 | 0,01 | 1,22 | 1,26 | 0,61 | 0,67 |
| PREL3 | -0,31 | 0,01 | 0,72 | 0,71 | 0,7 | 1,42 |
| PREL4 | 0,21 | 0,01 | 1,22 | 1,29 | 0,6 | 0,64 |
| VALU1 | 0,02 | 0,02 | 1,11 | 1,34 | 0,79 | 0,87 |
| VALU2 | -0,17 | 0,02 | 0,97 | 1,38 | 0,8 | 1 |
| VALU3 | 0,04 | 0,02 | 0,93 | 1,02 | 0,82 | 1,06 |
| VALU4 | 0,11 | 0,02 | 0,93 | 1,02 | 0,82 | 1,06 |
| ELEM1 | 0,07 | 0,02 | 0,89 | 0,92 | 0,7 | 1,18 |
| ELEM2 | 0,28 | 0,02 | 0,93 | 0,96 | 0,69 | 1,1 |
| ELEM3 | 0,3 | 0,02 | 1,19 | 1,28 | 0,61 | 0,73 |
| ELEM4 | -0,32 | 0,02 | 0,97 | 1,09 | 0,66 | 1,03 |
| ELEM5 | -0,32 | 0,02 | 1 | 1,13 | 0,64 | 0,99 |
| LOCU1 | -0,97 | 0,02 | 0,97 | 0,92 | 0,45 | 1,08 |
| LOCU2 | -0,87 | 0,02 | 1,06 | 1,12 | 0,41 | 0,93 |
| LOCU3 | 0,94 | 0,02 | 1,14 | 1,14 | 0,49 | 0,78 |
| LOCU4 | 0,11 | 0,02 | 0,81 | 0,8 | 0,61 | 1,37 |
| LOCU5 | 0,79 | 0,02 | 1,06 | 1,06 | 0,54 | 0,9 |
| ESTA1 | -1,2 | 0,02 | 1,1 | 1,2 | 0,34 | 0,87 |
| ESTA2 | -1,11 | 0,02 | 0,99 | 1,07 | 0,41 | 0,99 |
| ESTA3 | 0,49 | 0,02 | 0,88 | 0,89 | 0,57 | 1,2 |
| ESTA4 | 1,22 | 0,02 | 1,09 | 1,14 | 0,49 | 0,84 |
| ESTA5 | 0,59 | 0,02 | 0,92 | 0,95 | 0,55 | 1,15 |
| CONT1 | -1,26 | 0,02 | 1,22 | 1,26 | 0,37 | 0,78 |
| CONT2 | 0,77 | 0,02 | 1,07 | 1,07 | 0,57 | 0,9 |
| CONT3 | -0,12 | 0,02 | 0,77 | 0,78 | 0,65 | 1,4 |
| CONT4 | 0,61 | 0,02 | 1,01 | 1,01 | 0,6 | 1 |

Tabla 7.4: Estadísticos descriptivos para la faceta ítems para cada sub-escala

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| AUTE S1 | Media | 0,00 | 0,02 | 1,00 | 1,04 | 0,72 |
| | SD | 0,51 | 0,00 | 0,10 | 0,12 | 0,02 |
| | Q1 | -0,26 | 0,02 | 0,95 | 0,95 | 0,71 |
| | Q3 | 0,34 | 0,02 | 1,08 | 1,11 | 0,75 |
| ERES S2 | Media | 0,00 | 0,02 | 0,99 | 1,03 | 0,75 |
| | SD | 0,28 | 0,00 | 0,19 | 0,23 | 0,04 |
| | Q1 | -0,24 | 0,02 | 0,87 | 0,89 | 0,74 |
| | Q3 | 0,25 | 0,02 | 1,03 | 1,05 | 0,77 |
| MOTI S3 | Media | 0,00 | 0,02 | 0,98 | 1,13 | 0,81 |
| | SD | 0,38 | 0,00 | 0,09 | 0,10 | 0,01 |
| | Q1 | -0,30 | 0,02 | 0,91 | 1,04 | 0,81 |
| | Q3 | 0,29 | 0,02 | 1,02 | 1,18 | 0,81 |
| PREL S4 | Media | 0,00 | 0,01 | 1,00 | 1,05 | 0,65 |
| | SD | 0,26 | 0,00 | 0,23 | 0,24 | 0,04 |
| | Q1 | -0,23 | 0,01 | 0,80 | 0,88 | 0,61 |
| | Q3 | 0,24 | 0,01 | 1,22 | 1,27 | 0,69 |
| VALU S5 | Media | 0,00 | 0,02 | 0,99 | 1,19 | 0,81 |
| | SD | 0,10 | 0,00 | 0,07 | 0,17 | 0,01 |
| | Q1 | -0,03 | 0,02 | 0,93 | 1,02 | 0,80 |
| | Q3 | 0,06 | 0,02 | 1,01 | 1,35 | 0,82 |
| ELEM S6 | Media | 0,00 | 0,02 | 1,00 | 1,08 | 0,66 |
| | SD | 0,27 | 0,00 | 0,10 | 0,13 | 0,03 |
| | Q1 | -0,32 | 0,02 | 0,93 | 0,96 | 0,64 |
| | Q3 | 0,28 | 0,02 | 1,00 | 1,13 | 0,69 |
| LOCU S7 | Media | 0,00 | 0,02 | 1,01 | 1,01 | 0,50 |
| | SD | 0,80 | 0,00 | 0,11 | 0,13 | 0,07 |
| | Q1 | -0,87 | 0,02 | 0,97 | 0,92 | 0,45 |
| | Q3 | 0,79 | 0,02 | 1,06 | 1,12 | 0,54 |
| ESTA S8 | Media | 0,00 | 0,02 | 1,00 | 1,05 | 0,47 |
| | SD | 0,97 | 0,00 | 0,09 | 0,12 | 0,09 |
| | Q1 | -1,11 | 0,02 | 0,92 | 0,95 | 0,41 |
| | Q3 | 0,59 | 0,02 | 1,09 | 1,14 | 0,55 |
| CONT S9 | Media | 0,00 | 0,02 | 1,02 | 1,03 | 0,55 |
| | SD | 0,80 | 0,00 | 0,16 | 0,17 | 0,11 |
| | Q1 | -0,41 | 0,02 | 0,95 | 0,95 | 0,52 |
| | Q3 | 0,65 | 0,02 | 1,11 | 1,12 | 0,61 |

Tabla 7.5: Confiabilidad de Separación para la faceta Ítems para cada sub-escala.

| | <i>SR</i> | <i>G</i> | <i>H</i> | Prueba Efectos Fijos (todos iguales) valor p |
|----------|-----------|----------|----------|----------------------------------------------------------|
| AUTE(S1) | 1,00 | 31,21 | 41,94 | 0,00 |
| ERES(S2) | 0,99 | 13,80 | 18,73 | 0,00 |
| MOTI(S3) | 1,00 | 20,47 | 27,62 | 0,00 |
| PREL(S4) | 1,00 | 18,63 | 25,17 | 0,00 |
| VALU(S5) | 0,91 | 3,16 | 4,55 | 0,00 |
| ELEM(S6) | 0,99 | 13,07 | 17,76 | 0,00 |
| LOCU(S7) | 1,00 | 40,86 | 54,81 | 0,00 |
| ESTA(S8) | 1,00 | 50,62 | 67,83 | 0,00 |
| CONT(S9) | 1,00 | 40,49 | 54,32 | 0,00 |

Faceta Docentes.

En la Tabla 7.6 se muestran las distribuciones de las medidas y estadísticos de ajuste obtenidas para los docentes, para cada sub-escala. Las distribuciones que se encontraron de los valores de Infit y Outfit de los docentes, para todas las sub-escalas, indican un buen nivel de ajuste. No obstante, el ajuste encontrado en los docentes es menor que el ajuste encontrado en los ítems. Esto queda de manifiesto al comparar los valores medios y desvíos estándar encontrados para los índices de ajuste Infit y Outfit, entre la faceta docentes e ítems.

Por otra parte, el valor de la media y desvío estándar de r_{pm} , mostrado en la Tabla 7.6 indica un alto nivel de consistencia entre las medidas de los docentes y el constructo medido por la escala. Por otra parte, el nivel de consistencia encontrado es cercano al encontrado para los ítems.

El análisis de la confiabilidad de las medidas de los docentes se presenta en la Tabla 6.6. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 docentes tienen diferente efecto facilitador, para cada una de las sub-escalas consideradas.

Adicionalmente, se observan variaciones importantes en los valores de G entre las sub-escalas (G entre 1,30 (sub-escala ESTA S8) y 9,01 (sub-escala AUTE S1)). Estas variaciones, se aprecian también cuando se consideran los valores de confiabilidad de separación SR , o confiabilidad de Rasch, análogo de la confiabilidad de un test, en la teoría clásica de los test. Los valores de SR oscilaron entre 0,63 (sub-escala ESTA S8) y 0,99 (sub-escala AUTE S1). No obstante, los niveles de confiabilidad de las medidas de los docentes son buenos, salvo para la escala ESTA (S8) que es relativamente bajo.

Las variaciones encontradas en los niveles de confiabilidad se vieron reflejadas en el número de estratos estadísticamente discernibles (H) que podrían definirse para las medidas de los docentes. El valor de H para las medidas de los docentes osciló entre 2 (sub-escala ESTA S8) y cerca de 12 (sub-escala AUTE S1). Este resultado muestra que para todos los constructos considerados se encontraron al menos dos estratos estadísticamente discernibles del efecto facilitador del docente. Además, para 8 de 9 de los constructos medidos se encontraron un número considerable de estratos del efecto facilitador del docente (entre 4 y 12).

Tabla 7.6: Estadísticos descriptivos para la faceta Docentes para cada sub-escala.

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| AUTE S1 | Media | 0,00 | 0,09 | 1,16 | 1,22 | 0,70 |
| | SD | 0,90 | 0,04 | 0,38 | 0,49 | 0,10 |
| | Q1 | -0,64 | 0,06 | 0,84 | 0,83 | 0,65 |
| | Q3 | 0,67 | 0,11 | 1,43 | 1,44 | 0,77 |
| ERES S2 | Media | 0,00 | 0,16 | 1,14 | 1,20 | 0,69 |
| | SD | 1,05 | 0,08 | 0,39 | 0,51 | 0,13 |
| | Q1 | -0,77 | 0,12 | 0,79 | 0,78 | 0,63 |
| | Q3 | 0,81 | 0,20 | 1,48 | 1,49 | 0,79 |
| MOTI S3 | Media | 0,00 | 0,15 | 1,25 | 1,46 | 0,74 |
| | SD | 1,15 | 0,07 | 0,57 | 1,20 | 0,13 |
| | Q1 | -0,61 | 0,11 | 0,82 | 0,79 | 0,68 |
| | Q3 | 0,37 | 0,18 | 1,54 | 1,59 | 0,83 |
| PREL S4 | Media | 0,00 | 0,11 | 1,10 | 1,20 | 0,61 |
| | SD | 0,57 | 0,05 | 0,25 | 0,42 | 0,10 |
| | Q1 | -0,37 | 0,08 | 0,93 | 0,93 | 0,56 |
| | Q3 | 0,36 | 0,13 | 1,28 | 1,32 | 0,68 |
| VALU S5 | Media | 0,00 | 0,18 | 1,16 | 1,52 | 0,75 |
| | SD | 1,34 | 0,09 | 0,45 | 1,44 | 0,11 |
| | Q1 | -1,00 | 0,13 | 0,77 | 0,76 | 0,71 |
| | Q3 | 1,10 | 0,22 | 1,49 | 1,56 | 0,84 |
| ELEM S6 | Media | 0,00 | 0,16 | 1,03 | 1,17 | 0,63 |
| | SD | 0,58 | 0,08 | 0,32 | 0,53 | 0,12 |
| | Q1 | -0,27 | 0,11 | 0,79 | 0,79 | 0,56 |
| | Q3 | 0,35 | 0,20 | 1,27 | 1,43 | 0,72 |
| LOCU S7 | Media | 0,00 | 0,14 | 1,10 | 1,12 | 0,56 |
| | SD | 0,46 | 0,07 | 0,25 | 0,33 | 0,09 |
| | Q1 | -1,17 | 0,03 | 0,39 | 0,40 | 0,11 |
| | Q3 | 0,40 | 0,18 | 1,27 | 1,31 | 0,62 |
| ESTA S8 | Media | 0,00 | 0,14 | 1,08 | 1,16 | 0,59 |
| | SD | 0,25 | 0,07 | 0,22 | 0,31 | 0,07 |
| | Q1 | -0,17 | 0,10 | 0,92 | 0,93 | 0,56 |
| | Q3 | 0,13 | 0,17 | 1,18 | 1,33 | 0,64 |
| CONT S9 | Media | 0,00 | 0,16 | 1,11 | 1,17 | 0,58 |
| | SD | 0,53 | 0,08 | 0,28 | 0,38 | 0,11 |
| | Q1 | -0,47 | 0,11 | 0,91 | 0,90 | 0,54 |
| | Q3 | 0,46 | 0,20 | 1,29 | 1,37 | 0,64 |

Tabla 7.7: Confiabilidad de Separación para la faceta Docentes para cada sub-escala.

| | SR | G | H | Prueba Efectos Fijos (todos iguales) valor p |
|----------|------|------|-------|----------------------------------------------------------|
| AUTE(S1) | 0,99 | 9,01 | 12,35 | 0,00 |
| ERES(S2) | 0,97 | 5,68 | 7,90 | 0,00 |
| MOTI(S3) | 0,98 | 6,66 | 9,22 | 0,00 |
| PREL(S4) | 0,95 | 4,42 | 6,22 | 0,00 |
| VALU(S5) | 0,98 | 6,59 | 9,12 | 0,00 |
| ELEM(S6) | 0,91 | 3,15 | 4,53 | 0,00 |
| LOCU(S7) | 0,89 | 2,77 | 4,03 | 0,00 |
| ESTA(S8) | 0,63 | 1,30 | 2,07 | 0,00 |
| CONT(S9) | 0,89 | 2,80 | 4,07 | 0,00 |

Faceta Estudiantes.

En la Tabla 7.8 se muestran las distribuciones de las medidas y estadísticos de ajuste obtenidas para los estudiantes, para cada sub-escala. Las distribuciones que se encontraron de los valores de Infit y Outfit de los estudiantes, para todas las sub-escalas, indican un nivel de ajuste aceptable. No obstante, el ajuste encontrado en los estudiantes es menor que el ajuste encontrado para los ítems y docentes. Esto queda en evidencia al comparar los valores medios y desvíos estándar encontrados para los índices de ajuste Infit y Outfit, entre la faceta estudiantes, y las facetas docentes e ítems. En especial, si bien los valores medios de Infit y Outfit son muy próximos al valor esperado para un ajuste perfecto entre los datos y el modelo, los desvíos estándar de ambos estadísticos son mucho mayores que para las otras dos facetas.

Por otra parte, la distribución de los valores de r_{pm} para cada una de las sub-escalas, mostrada en la Tabla 7.8 indica que en general se encontró un nivel de consistencia aceptable entre las medidas de los estudiantes y el constructo medido por la sub-escala. Para todas las sub-escalas se encuentra que al menos un 75% de los valores encontrados de r_{pm} son no negativos. Por otra parte, el nivel de consistencia encontrado para las medidas de los estudiantes es mucho menor que la observada para los ítems y docentes. Esto se aprecia claramente al comparar la distribución de r_{pm} para los estudiantes con la distribución de r_{pm} encontrada en las otras facetas.

El análisis de la confiabilidad de las medidas de los estudiantes se presenta en la Tabla 7.9. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos 2 estudiantes tienen diferente medida, para cada una de las sub-escalas consideradas.

Los valores de los índices de confiabilidad muestran que en general la confiabilidad de las medidas es aceptable. La escala ESTA (S8) presentó el valor más bajo de confiabilidad, no obstante permite definir dos estratos de estudiantes de acuerdo la dimensión de estabilidad de las atribuciones causales. Por otra parte las escalas AUTE(S1), ERES(S2), MOTI(S3) y VALU(S5) tienen confiabilidades entre buenas y elevadas. Estas sub-escalas permiten definir alrededor de 4 a 5 estratos.

Tabla 7.8: Estadísticos descriptivos para la faceta Estudiantes para cada sub-escala.

| | | Medida | S.E. | Infit | Outfit | r_{pm} |
|------------|-------|--------|------|-------|--------|----------|
| AUTE S1 | Media | 1,10 | 0,73 | 1,01 | 1,02 | 0,18 |
| | SD | 2,26 | 0,36 | 0,93 | 0,97 | 0,45 |
| | Q1 | -0,48 | 0,45 | 0,39 | 0,39 | 0,00 |
| | Q3 | 2,53 | 0,86 | 1,34 | 1,42 | 0,56 |
| ERES S2 | Media | 1,10 | 0,73 | 1,01 | 1,02 | 0,18 |
| | SD | 2,26 | 0,36 | 0,93 | 0,97 | 0,45 |
| | Q1 | -0,48 | 0,45 | 0,39 | 0,39 | 0,00 |
| | Q3 | 2,53 | 0,86 | 1,34 | 1,42 | 0,56 |
| MOTI S3 | Media | 1,39 | 0,71 | 0,94 | 0,98 | 0,27 |
| | SD | 2,68 | 0,38 | 0,98 | 1,09 | 0,47 |
| | Q1 | -0,29 | 0,42 | 0,32 | 0,32 | 0,00 |
| | Q3 | 3,05 | 0,86 | 1,19 | 1,22 | 0,73 |
| PREL S4 | Media | 0,34 | 0,59 | 1,02 | 1,03 | 0,23 |
| | SD | 1,52 | 0,42 | 0,68 | 0,72 | 0,45 |
| | Q1 | -0,59 | 0,32 | 0,57 | 0,54 | 0,00 |
| | Q3 | 1,05 | 0,62 | 1,38 | 1,33 | 0,60 |
| VALU S5 | Media | 2,56 | 0,80 | 0,96 | 1,01 | 0,15 |
| | SD | 2,80 | 0,41 | 1,11 | 1,25 | 0,48 |
| | Q1 | 0,54 | 0,47 | 0,24 | 0,23 | 0,00 |
| | Q3 | 4,23 | 0,94 | 1,05 | 1,07 | 0,49 |
| ELEM S6 | Media | 1,24 | 0,78 | 0,99 | 1,01 | 0,17 |
| | SD | 1,95 | 0,49 | 0,64 | 0,72 | 0,39 |
| | Q1 | 0,06 | 0,43 | 0,61 | 0,60 | 0,00 |
| | Q3 | 2,38 | 0,86 | 1,32 | 1,37 | 0,51 |
| LOCU S7 | Media | 0,87 | 0,64 | 1,05 | 1,05 | 0,36 |
| | SD | 1,36 | 0,33 | 0,70 | 0,79 | 0,43 |
| | Q1 | -0,15 | 0,40 | 0,55 | 0,48 | 0,00 |
| | Q3 | 1,62 | 0,78 | 1,33 | 1,28 | 0,70 |
| ESTA S8 | Media | 0,58 | 0,62 | 1,03 | 1,06 | 0,44 |
| | SD | 1,22 | 0,31 | 0,58 | 0,75 | 0,44 |
| | Q1 | -0,03 | 0,37 | 0,60 | 0,57 | 0,00 |
| | Q3 | 1,11 | 0,74 | 1,33 | 1,28 | 0,79 |
| CONT S9 | Media | 0,70 | 0,73 | 1,04 | 1,05 | 0,37 |
| | SD | 1,52 | 0,36 | 0,71 | 0,91 | 0,42 |
| | Q1 | -0,38 | 0,45 | 0,55 | 0,43 | 0,00 |
| | Q3 | 1,58 | 0,90 | 1,35 | 1,35 | 0,73 |

Tabla 7.9: Confiabilidad de Separación para la faceta Estudiantes para cada sub-escala.

| | SR | G | H | Prueba Efectos Fijos (todos iguales) valor p |
|----------|------|------|------|----------------------------------------------------------|
| AUTE(S1) | 0,93 | 3,55 | 5,07 | 0,00 |
| ERES(S2) | 0,87 | 2,60 | 3,80 | 0,00 |
| MOTI(S3) | 0,91 | 3,17 | 4,56 | 0,00 |
| PREL(S4) | 0,77 | 1,85 | 2,81 | 0,00 |
| VALU(S5) | 0,90 | 2,95 | 4,27 | 0,00 |
| ELEM(S6) | 0,77 | 1,85 | 2,80 | 0,00 |
| LOCU(S7) | 0,72 | 1,60 | 2,46 | 0,00 |
| ESTA(S8) | 0,68 | 1,45 | 2,27 | 0,00 |
| CONT(S9) | 0,71 | 1,58 | 2,45 | 0,00 |

7.3.3. Invarianza.

La invarianza de las medidas de los estudiantes y docentes respecto de los ítems significa que las medidas de los primeros no dependen de la muestra específica de ítems considerados. Una elevada invarianza respecto a los ítems se corresponde con niveles elevados de correlación lineal entre las medidas obtenidas con diferentes submuestra de ítems. De esta forma, se espera una correlación lineal elevada entre las medidas obtenidas para sub-muestras aleatorias complementarias, así como también correlaciones lineales elevadas entre cada sub-muestra y la muestra completa. En la Tabla 7.10 se muestran los resultados del análisis de invarianza de las medidas de elementos de las facetas estudiantes y docentes, respecto a los ítems.

Tabla 7.10: Análisis de la Invarianza de las Medidas para cada sub-escala

| Sub-escala | Correlación entre muestras | Invarianza Respecto a Ítems | |
|------------|----------------------------|-----------------------------|---------------|
| | | Estudiantes | Docentes |
| AUTE (S1) | $r(1, 2)$ | 0,91 (1,00) | 0,96 (1,00) |
| | $r(1, C)$ | 0,98 (1,00) | 0,99 (1,00) |
| | $r(2, C)$ | 0,96 (1,00) | 0,99 (1,00) |
| ERES (S2) | $r(1, 2)$ | 0,75 (0,92) | 0,58 (0,61) |
| | $r(1, C)$ | 0,92 (1,00) | 0,67 (0,69) |
| | $r(2, C)$ | 0,93 (1,00) | 0,96 (1,00) |
| MOTI (S3) | $r(1, 2)$ | 0,84 (0,97) | 0,87 (0,90) |
| | $r(1, C)$ | 0,96 (1,00) | 0,96 (0,99) |
| | $r(2, C)$ | 0,95 (1,00) | 0,97 (0,99) |
| PREL (S4) | $r(1, 2)$ | 0,54 (0,74) | 0,46 (0,49) |
| | $r(1, C)$ | 0,84 (1,00) | 0,76 (0,81) |
| | $r(2, C)$ | 0,87 (1,00) | 0,49 (0,51) |
| VALU (S5) | $r(1, 2)$ | 0,85 (1,00) | 0,95 (0,99) |
| | $r(1, C)$ | 0,96 (1,00) | 0,99 (1,00) |
| | $r(2, C)$ | 0,96 (1,00) | 0,99 (1,00) |
| ELEM (S6) | $r(1, 2)$ | 0,65 (0,95) | -0,13 (-0,14) |
| | $r(1, C)$ | 0,93 (1,00) | 0,55 (0,59) |
| | $r(2, C)$ | 0,85 (1,00) | 0,75 (0,81) |
| LOCU (S7) | $r(1, 2)$ | 0,35 (0,53) | 0,43 (0,49) |
| | $r(1, C)$ | 0,83 (1,00) | 0,71 (0,79) |
| | $r(2, C)$ | 0,78 (1,00) | 0,89 (1,00) |
| ESTA (S8) | $r(1, 2)$ | 0,46 (0,96) | 0,28 (0,52) |
| | $r(1, C)$ | 0,90 (1,00) | 0,87 (1,00) |
| | $r(2, C)$ | 0,78 (1,00) | 0,71 (1,00) |
| CONT (S9) | $r(1, 2)$ | 0,43 (0,88) | 0,55 (1,00) |
| | $r(1, C)$ | 0,79 (1,00) | 0,87 (1,00) |
| | $r(2, C)$ | 0,87 (1,00) | 0,83 (1,00) |

Cuando se comparan las invarianzas entre la faceta de estudiantes y la de docentes, se aprecia que los valores obtenidos de los coeficientes de correlación lineal sugieren que, en términos generales, la invarianza en la faceta estudiantes es algo mejor que en la faceta docentes. En particular, uno de los casos más extremos se verifica para la escala ELEM (S6). En este caso, las medidas de los

docentes presentaron una pobre invarianza, mientras que para los estudiantes alcanza un nivel bueno. Por otra parte, algunas de las escalas se destacan por permitir obtener medidas con un alto nivel de invarianza para las dos facetas consideradas (AUTE(S1) y VALU(S5)).

7.3.4. Interacciones (sesgos) entre los elementos de las facetas.

En esta sección se analizan las interacciones entre las tres facetas de medición (docentes, estudiantes e ítems). En la Tabla 7.11 se resumen los resultados obtenidos.

Tabla 7.11: Análisis de Interacciones (sesgos) entre facetas para cada dimensión considerada. En la Tabla: D (docente), E (Estudiante), I (Ítem), * (interacción), Var. Exp (%) (porcentaje de la varianza explicada por la interacción).

| | D*I | | D*E | | E*I | |
|----------|---------------|--------------------------------|---------------|--------------------------------|---------------|--------------------------------|
| | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) | Var. Exp. (%) | Prueba Efectos Fijos (valor p) |
| AUTE(S1) | 0,76 | 0,00 | 10,70 | 0,00 | 24,02 | 0,00 |
| ERES(S2) | 1,70 | 0,00 | 8,89 | 1,00 | 27,59 | 0,00 |
| MOTI(S3) | 0,42 | 0,21 | 11,85 | 0,00 | 16,11 | 0,00 |
| PREL(S4) | 1,53 | 0,00 | 8,32 | 1,00 | 16,11 | 0,00 |
| VALU(S5) | 0,93 | 0,00 | 11,19 | 0,99 | 19,36 | 0,00 |
| ELEM(S6) | 2,05 | 0,00 | 9,70 | 1,00 | 35,64 | 0,04 |
| LOCU(S7) | 1,74 | 0,00 | 4,91 | 1,00 | 41,65 | 0,00 |
| ESTA(S8) | 1,55 | 0,00 | 5,13 | 1,00 | 39,74 | 0,00 |
| CONT(S9) | 1,79 | 0,00 | 6,23 | 1,00 | 37,99 | 0,02 |

Un patrón claro emerge de los resultados de los análisis de las interacciones. En primer lugar, las interacciones entre la faceta estudiantes e ítems (Estudiante*Ítem) son estadísticamente significativas en todos los casos y explican un porcentaje elevado de la variabilidad observada. Esto implica que cuando se incorpora en el modelo estos términos de interacción el PVE o varianza explicada alcanza en todos los casos un nivel entorno a un 80%. El significado de este resultado se puede ilustrar considerando la interacción más fuerte que correponde a la interacción Estudiante*Ítem para la escala LOCU(S7).

Este término de interacción significa que, desde la perspectiva de un estudiante con un locus interno de causalidad, la dificultad de los ítems es percibida

de forma muy diferente a la dificultad percibida por estudiantes con locus de causalidad externo. Por tanto, los estudiantes con diferencias en el locus de causalidad tienden a adherir o acordar con los ítems en formas muy diferentes. Esto, podría reflejar una forma distinta de interpretar los enunciados de los ítems, mediada cognitivamente por las diferencias en el locus de control.

En segundo lugar, las interacciones Docente*Ítem en general fueron estadísticamente significativas, pero, en todos los casos la varianza explicada por el término de interacción es muy pequeño. Así, el incorporar dicho término de interacción al modelo con efectos primarios (aditivos), prácticamente no incrementa el porcentaje de la varianza explicada por el modelo. El ejemplo más extremo de esto se da para la escala MOTI(S3) donde el porcentaje de la varianza explicada por el modelo pasa de 67 % a 67,42 %.

En tercer lugar, se encontró un grupo de interacciones, en general no estadísticamente significativas, que corresponden a interacciones Docente*Estudiante, que explican un porcentaje de la varianza entre un 4 % y 12 %. Debido a que estas interacciones no son significativas, no se puede descartar que, una vez introducido el error de medición, los términos de interacción Docente*Estudiante tengan todos medida cero, y los valores observados se deban a variaciones aleatorias relacionadas con errores de medición. Las excepciones, en este grupo de las interacciones Docente*Estudiante, son: el término de interacción correspondiente a la sub-escala AUTE (S1) y el correspondiente a la escala MOTI (S3). Estas interacciones son altamente significativas y explican alrededor de un 11 % de la varianza observada.

La sub-escala AUTE(S1) esta relacionada con el constructo auto-eficacia para aprender, mientras que, la sub-escala MOTI(S3) está relacionada con el constructo motivación intrínseca. Las interacciones encontradas significan que algunos estudiantes pueden experimentar un nivel de auto-eficacia para aprender y una motivación intrínseca, que no puede ser solo explicada por la un efecto aditivo de valores estables del estudiantes y el efecto facilitador del docente.

Una posible explicación para esto, podría ser que el efecto facilitador del docente sobre la auto-eficacia para aprender y la motivación intrínseca, no afecta de igual forma a estudiantes con bajos niveles de motivación intrínseca y auto-eficacia para aprender, que a estudiantes con altos niveles. Para que esto explique el patrón encontrado de interacciones, además, las diferencias mencionadas, no deberían ocurrir en igual grado con los otros constructos

estudiados, que no presentaron interacciones estadísticamente significativas.

7.4. Discusión.

La investigación reportada en este capítulo comprendió dos objetivos principales. Un primer objetivo fue aportar evidencias de validez del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. Un segundo objetivo fue aportar nuevas evidencias que permitan evaluar la hipótesis central de esta investigación.

La hipótesis central considerada en esta tesis expresa que existen diferencias individuales en cómo los estudiantes evalúan la enseñanza que están relacionadas con variables que median la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza, o bien, que explican diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes. En particular, en esta investigación se consideraron las siguientes variables: auto-eficacia para aprender, expectativas positivas de resultados, motivación intrínseca, percepción de elección, valor/utilidad, elección de metas y dimensiones de atribuciones causales, es decir, locus de control, estabilidad y controlabilidad.

En la investigación reportada en el capítulo 6 se argumentó que, como una consecuencia lógica de la hipótesis de investigación planteada, se esperaría un menor ajuste de la faceta estudiantes, en relación al modelo MFRM sobre el proceso de respuesta a cuestionarios de evaluación de la calidad de la enseñanza. Esto, estaría relacionado con cambios en la severidad del estudiante que serían resultado de la interacción entre características del estudiante y las características de la enseñanza. El patrón de ajuste predicho y las interacciones propuestas fueron respaldadas por las evidencias.

Por ejemplo, supongamos que la severidad del estudiante al evaluar la enseñanza está relacionada con la motivación intrínseca experimentada por el estudiante. En este trabajo asumimos que la motivación intrínseca tendría un componente estable (MI_{Est}) que depende de características del estudiante y un componente que corresponde al efecto facilitador del docente (MI_{Doc}). El efecto facilitador del docente sobre la motivación intrínseca se conceptualiza también como una característica estable del docente. Si el efecto facilitador sobre la motivación intrínseca presenta cierta variación entre los docentes, entonces, la motivación intrínseca experimentada por el estudiante (MI_{Exp}) variaría según la siguiente relación $MI_{Exp} = MI_{Est} + MI_{Doc} + MI_{Est} * MI_{Doc}$ entre

actividades de enseñanza.

El menor ajuste de los estudiantes en relación al modelo MFRM considerado en el capítulo 6 podría ser una consecuencia de que los parámetros de severidad (severidad “global”) reflejan el valor MI_{Est} , mientras que la severidad del estudiante al evaluar cierto docente en particular (severidad “local”) esta relacionada con MI_{Exp} . Esto tendría como consecuencia que la severidad “local” sería distinta de la severidad “global”, lo que se manifestaría como una interacción Docente*Estudiante.

Una consecuencia lógica de los argumentos anteriores es que la severidad global de diferentes estudiantes estaría relacionada con MI_{Est} . Por otra parte, las diferencias en la severidad local de un estudiante respecto de diferentes docentes estaría también relacionada con las diferencias en el efecto facilitador de los docentes, y eventualmente, de posibles interacciones Docente*Estudiante.

De acuerdo a los resultados reportados en el capítulo 6: existe una gran variabilidad en la severidad global de los estudiantes al evaluar la calidad de la enseñanza; en segundo lugar, para todas las dimensiones consideradas, existen diferencias en la severidad local de los estudiantes, que se manifestaron como interacciones Docentes*Estudiante.

A partir de estos resultados, dos predicciones teóricas son una consecuencia lógica de este patrón y de la explicación del proceso de respuesta deducida a partir de la hipótesis central de esta investigación: en primer lugar, en la muestra de calibración estudiada debería existir una considerable variabilidad entre los estudiantes en los constructos medidos (por ejemplo, en MI_{Est}); en segundo lugar, debería existir variabilidad en el efecto facilitador de los docentes en relación a cada uno de los constructos considerados.

Ambas predicciones teóricas son ampliamente consistentes con los resultados obtenidos. En relación a las medidas obtenidas para los estudiantes a partir de todas las sub-escalas se encontró que existen varios estratos estadísticamente discernibles, como lo indican los resultados de la prueba de efectos fijos (todos iguales) y los valores encontrado del estadístico H. Esto se verificó también, en forma clara, para el efecto facilitador de los docentes.

Observar que la discusión previa vinculada con la hipótesis central de esta investigación constituye en si misma un análisis de validez de constructo fundamentada en el estudio teórico y empírico del proceso de respuesta desde una perspectiva cognitiva y socio-cognitiva. Esta perspectiva representa un abordaje a la validez de constructo con algunas similitudes al análisis de re-

presentación de constructo (Embretson, 1994) y al enfoque substantivo a la validez de constructo (Messick, 1995).

Algunas otras consideraciones vinculadas con los resultados obtenidos en esta investigación aportan elementos adicionales para discutir la validez de las escalas consideradas.

En primer lugar, los resultados de ajuste entre los datos y el modelo a nivel global y a nivel de cada faceta son consistentes con un acuerdo razonable de las mediciones obtenidas con los criterios ideales y bastante exigentes implícitos en los modelo de múltiples facetas de Rasch. En relación al ajuste de las facetas, se encontró un buen ajuste en general. En particular el ajuste de los ítems fue muy elevado. Además, los valores encontrados de r_{pm} muestran un nivel alto de consistencia entre las medidas de los ítems y el constructo medido por cada sub-escala. Además, los niveles de confiabilidad de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) muestra que los estudiantes son capaces de apreciar las diferencias que existen de contenido entre los ítems.

Adicionalmente los análisis de la invarianza de las medidas mostraron que varias de las escalas consideradas presentan un nivel elevado de la invarianza de las medidas de los estudiantes y docentes respecto de los ítems. No obstante, los niveles de invarianza para las medidas de las escalas presentaron algunas variaciones. En particular, la escala ELEM (S6) mostró un nivel bajo de invarianza, que puede ser útil investigar su origen tanto desde un punto de vista aplicado como teórico.

Por otra parte, en general los valores de dificultad de los ítems mostraron una distribución razonable a lo largo de la amplitud de variación de los constructos considerados (Tabla 7.3). En general, las medidas de dificultad de los ítems resultaron consistentes con expectativas teóricas que surgen de análisis del significado del constructo y de los enunciados de los ítems. Esto fue especialmente claro en los análisis realizados de la escala ATRI. Las magnitudes de las dificultades de los ítems para LOCU(S7), ESTA(S8) y CONT(S9), acuerdan claramente con expectativas fundadas en la teoría de las atribuciones causales (Weiner, 1983, 2013).

Más allá de las consideraciones teóricas realizadas este trabajo tiene importantes implicaciones prácticas. El presente trabajo muestra la factibilidad de una aproximación complementaria a la evaluación de la calidad de la enseñanza de los docentes. Esta aproximación se basa en la medición y evaluación

diagnóstica tanto del proceso de aprendizaje de los estudiantes como del efecto facilitador de los docentes. Esta aproximación puede permitir la implementación de evaluaciones diagnósticas y formativas dirigidas tanto a estudiantes como docentes.

En conclusión, los resultados encontrados aportaron evidencias de la validez del cuestionario considerado y son consistentes con las predicciones realizadas.



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Capítulo 8

Modelado del proceso de respuesta a cuestionarios de evaluación estudiantil de la calidad en la enseñanza aplicando teorías socio-cognitivas de la motivación.

8.1. Presentación.

En este capítulo se reporta una investigación realizada con el propósito de proveer una evaluación más directa de la hipótesis central de esta tesis, así como también, aportar nuevas evidencias de la validez de constructo del cuestionario de evaluación de la calidad en la enseñanza y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje.

Un aspecto relevante a los objetivos planteados, está vinculado con la relación entre la severidad global de los estudiantes y las medidas de los estudiantes en nueve variables socio-cognitivas consideradas en esta tesis: auto-eficacia para aprender, expectativas positivas de resultado, motivación intrínseca, percepción de elección, percepción de valor y utilidad, elección de metas, locus de

causalidad, estabilidad y controlabilidad.

La aproximación metodológica se fundamentó en la aplicación de diferentes abordajes metodológicos al mismo objeto de estudio. En primer lugar, se aplicó un modelo de medidas invariantes explicativo (LLT-MFRM) que permitió descomponer la medida de severidad global en una serie de componentes o factores de severidad (Engelhard Jr, 2013; M. Wilson y De Boeck, 2004).

De acuerdo a la hipótesis principal, se espera que para cada dimensión de la enseñanza, la medida de severidad presente al menos un componente con medida no nula. En forma similar, se espera que cada componente este relacionado con al menos una medida de severidad de las dimensiones de la enseñanza. El peso exacto de cada componente se espera este relacionado con el procesamiento de información cognitiva vinculada a la evaluación de cada dimensión de la enseñanza en particular y el papel de mediación cognitiva, que presentan las variables socio-cognitivas consideradas.

En segundo lugar, se realizó un análisis correlacional entre las medidas de calidad en la enseñanza y las medidas de las variables socio-cognitivas, para analizar las evidencias de validez convergente y discriminante (Albright y Malloy, 2000; Campbell y Fiske, 1959).

A partir de la hipótesis central de esta tesis se predice que debería existir una correlación estadísticamente significativa entre el primer conjunto de variables y el segundo. Adicionalmente, de acuerdo a los modelos teóricos del proceso de respuesta propuestos en el capítulo 4, se espera una correlación negativa entre las medidas de severidad y las medidas de las variables socio-cognitivas (véase capítulo 4).

En tercer lugar, se derivaron modelos de regresión para la variable severidad global, en los que la selección de las variables explicativas se fundamentó en los modelos teóricos propuestos en el capítulo 4. A partir de estos modelos se predijo el patrón esperado de efectos en los modelos de regresión relacionados con cada modelo teórico.

Los resultados obtenidos son ampliamente consistentes con las predicciones teóricas realizadas y constituyen una evidencia fuerte para la hipótesis central de esta investigación. Adicionalmente, los resultados obtenidos triangulando los métodos y modelos teóricos considerados, señalaron consistentemente que las variables socio-cognitivas consideradas explican diferencias en la severidad de los estudiantes al evaluar la enseñanza.

Adicionalmente, se encontraron evidencias importantes de validez de cons-

tructo del cuestionario de evaluación de calidad en la enseñanza y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje.

También, los resultados obtenidos muestran claramente que las variables socio-cognitivas consideradas serían importantes variables explicativas de la severidad global de los estudiantes al evaluar la calidad de la enseñanza. Además, en general tendrían un efecto negativo sobre la severidad global de los estudiantes.

8.2. Metodología.

La metodología básica empleada en este capítulo fue descrita previamente en el capítulo 3 y el capítulo 5 sección 5.2. En esta sección se describen aspectos específicos correspondientes al problema de investigación abordado.

8.2.1. Muestra de datos de calibración.

La muestra de datos de calibración consistió en las respuestas de 3243 estudiantes: a nueve sub-escalas del cuestionario de evaluación estudiantil de la calidad en la enseñanza; y, a siete sub-escalas del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. Las respuestas se obtuvieron durante dos semestres consecutivos, en 10 actividades de enseñanza, en las que participaron 98 docentes, en la carrera de médico durante el año 2015.

8.2.2. Instrumentos.

En la investigación reportada en este capítulo se consideró en forma conjunta el cuestionario de evaluación de la calidad en la enseñanza y el cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. Los instrumentos considerados fueron detalladamente descritos en los capítulos 3, 6 y 7. Las propiedades psicométricas y las evidencias de validez del cuestionario de evaluación de la calidad en la enseñanza se reportaron en el capítulo 6 de esta tesis. Por su parte, el estudio de las propiedades psicométricas y validez del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje fueron consideradas en el capítulo 7.

8.2.3. Aproximación Analítica.

La aproximación analítica se basó en dos enfoques complementarios. En primer lugar, las respuestas de los estudiantes a cada sub-escala del cuestionario de evaluación de la calidad de la enseñanza fue analizada por separado aplicando un modelo de múltiples facetas de Rasch con restricciones lineales en los parámetros de severidad (LLT-MFRM). En el modelo LLT-MFRM la faceta estudiantes fue definida como una faceta dummy con medida cero. Por otra parte, se introdujo una nueva faceta que contiene como elementos los componentes o factores de severidad. Así, en lugar de estimar la medida de los estudiantes, se determina la medida de los componentes o factores de severidad. En este modelo la severidad de un estudiante corresponde a la suma ponderada de las medidas de los factores de severidad.

Por tanto, el modelo LLT-MFRM utilizado para analizar los datos de respuestas a escalas de evaluación de la efectividad en la enseñanza puede ser especificado como sigue:

$$\ln \left[\frac{P_{nijk}}{P_{nijk-1}} \right] = \theta_n - \beta_i - \alpha_j - \tau_{ik} \quad (8.1)$$

$$\alpha_j = \sum_{l=0}^{l=p} w_{jl} * \eta_l \quad (8.2)$$

donde

P_{nijk} = Probabilidad de que el docente n reciba un puntaje k en el ítem i del estudiante j ,

P_{nijk-1} = Probabilidad de que el docente n reciba un puntaje $k-1$ en el ítem i del estudiante j ,

θ_n = Habilidad del docente n en la dimensión considerada,

β_i = Dificultad del ítem i ,

α_j = Severidad del estudiante j ,

τ_{ik} = Dificultad de recibir un puntaje k relativo a recibir un puntaje $k-1$ para el ítem i .

η_l = Severidad del componente l .

w_{jl} = Peso del componente l en el estudiante j .

El modelo LLT-MFRM puede considerarse una extensión del modelo logístico lineal de rasgo latente o LLTM (Fischer, 1977). El modelo LLTM

es un modelo de Rasch dicotómico con restricciones lineales en los parámetros de dificultad de los ítems. El modelo LLT-MFRM podría considerarse una generalización del modelo LLTM, para múltiples facetas, en las que se considera la mediación de evaluadores y restricciones lineales en los parámetros de severidad. El modelo LLT-MFRM es un modelo de respuesta a ítems explicativo (M. Wilson y De Boeck, 2004).

Tabla 8.1: Descripción de los componentes de Severidad en el modelo LLT-MFRM

| Componentes de Severidad | Variable Relacionada (sub-escala) | H para la medida de la variable | Pesos para el componente de Severidad (w_{ji}) |
|--------------------------|-------------------------------------------------|---------------------------------|----------------------------------------------------|
| AUTEE[C1] | Auto-eficacia para aprender (AUTE[S1]) | 5,07 | 1,2,3,4,5 |
| ERESE[C2] | Expectativas Positivas de Resultados (ERES[S2]) | 3,80 | 1,2,3,4,5 |
| MOTIE[C3] | Motivación Intrínseca (MOTI[S3]) | 4,56 | 1,2,3,4,5 |
| PRELE[C4] | Percepción de Elección (PREL [S4]) | 2,81 | 1,2,3 |
| VALUE[C5] | Valor/Utilidad (VALU[S5]) | 4,27 | 1,2,3,4,5 |
| ELEME[C6] | Elección de Metas (ELEM[S6]) | 2,80 | 1,2,3 |
| LOCUE[C7] | Locus de Causalidad (LOCU[S7]) | 2,46 | 1,2,3 |
| ESTAE[C8] | Estabilidad (ESTA[S8]) | 2,27 | 1,2,3 |
| CONTE[C9] | Controlabilidad (CONT[S9]) | 2,45 | 1,2,3 |

Los componentes o factores de severidad definidos en el modelo y sus pesos posibles se describen en la Tabla 8.1. Los pesos w_{ji} de cada componente fueron definidos como niveles de la variable relacionada con el componente. El número de niveles definido fue 3 para variables con un número de estratos estadísticamente discernibles próximo a 3 (con H entre 2,27 y 2,81) y 5 para las restantes variables (con H entre 3,80 y 5,07). Así, la definición de los componentes o factores de severidad se basó en las medidas invariantes de los estudiantes obtenidas a partir del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje.

El modelo LLT-MFRM fue aplicado para obtener una evaluación general de la hipótesis central de esta investigación. A partir de esta hipótesis se espera que las medidas de los componentes o factores de severidad sean distintos de cero y que el ajuste global entre los datos y el modelo sea aceptable. Además, se espera un ajuste aceptable para los componentes de la faceta de factores de severidad.

Debido a que el modelo LLT-MFRM contiene solo nueve elementos de medición (los nueve componentes de la faceta de factores de severidad) que substituyen 3243 elementos de la faceta estudiantes en el modelo MFRM original, se espera un ajuste mucho menor. Además, los factores definidos no son

un conjunto exhaustivo de los componentes involucrados en el proceso de respuesta, lo que implicaría también un menor ajuste esperado entre los datos y el modelo.

Adicionalmente, la definición de los pesos en el modelo es aproximada. Fundamentalmente los pesos intentan reflejar aspectos esenciales de la hipótesis propuesta. Así, para que una definición más detallada de los pesos sea posible, será necesario investigación adicional. Esto podría implicar mejoras en la precisión de la caracterización del proceso de respuesta.

No obstante, un análisis sistemático de los siguientes puntos se considera informativo en relación a la validez de la hipótesis de investigación propuesta: (1) el patrón de ajuste global del modelo para las nueve dimensiones de calidad en la enseñanza consideradas; (2) el patrón de ajuste de los componentes para cada una de las dimensiones; y (3), los patrones de las medidas de los componentes para cada dimensión.

Un segundo enfoque en la aproximación analítica consistió en un análisis en dos etapas. La primer etapa, de medición, estimó las medidas invariantes de los estudiantes para cada una de las dimensiones en los dos instrumentos considerados mediante análisis de múltiples facetas de Rasch implementado con el programa Facets (Linacre, 2015). Esta etapa de medición, fue seguida por una etapa de análisis de la relación entre las medidas de severidad de los estudiantes y las medidas obtenidas a partir del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. Este tipo de enfoque, introducido por Scheiblechner (1972), para tratar un problema similar al abordado con modelos LLTM (Fischer, 1977), combina modelos de respuesta a los ítems y modelos de regresión lineal. Este enfoque representa una aproximación más flexible y conveniente, que permite obtener resultados similares a enfoques exclusivamente basados en modelos de teoría de respuesta a los ítems (Embretson y Reise, 2000; Green y Smith, 1987). La etapa de análisis fue implementada usando R versión 3.4.3 (R Development Core Team, 2008).

Esta segunda etapa comprendió en primer lugar el análisis del patrón de correlaciones entre las medidas de severidad y las medidas de las variables socio-cognitivas. A partir de la hipótesis central de esta tesis se predice que debería existir una correlación estadísticamente significativa entre el primer conjunto de variables y el segundo. Adicionalmente, de acuerdo a los modelos teóricos del proceso de respuesta propuestos en el capítulo 4, se espera una correlación negativa entre las medidas de severidad y las medidas de las variables

socio-cognitivas (véase capítulo 4).

Así, correlaciones estadísticamente significativas entre los dos conjuntos de variables consideradas son consistentes con la hipótesis de investigación, mientras que, correlaciones estadísticamente significativas, negativas, adicionalmente son consistentes con una predicción común de los modelos teóricos propuestos. Estos modelos representan especificaciones de hipótesis de validez de las escalas consideradas, derivadas de la hipótesis central de esta investigación y de las diferentes perspectivas teóricas abordadas. En el análisis de correlaciones se realizaron correcciones por atenuación y se calcularon valores *p* corregidos para controlar el error de familia o experimental (para inferencias simultáneas) usando el método de Holm (Holm, 1979) utilizando la función `p.adjust` del paquete `stats`.

Un aspecto importante de esta aproximación metodológica es que permite conocer la relación entre las dimensiones de calidad en la enseñanza y cada variable socio-cognitiva, sin restringir ni alterar la natural variación y covariación de las variables consideradas. Esto es consistente con una aproximación de diseño representativo (Albright y Malloy, 2001; Brunswik, 1947, 1955). Esto permite apreciar las relaciones entre los constructos considerados manteniendo el patrón multivariado natural de covariación de observaciones potenciales (Brunswik, 1956; Hammond y Stewart, 2001).

En segundo lugar, en la etapa del análisis de este último enfoque, se derivaron modelos de regresión para la variable respuesta severidad global, en los que la selección de las variables explicativas se fundamentó en los modelos teóricos propuestos en el capítulo 4. En la Tabla 8.2 se describe la relación entre los cuatro modelos teóricos, los modelos estadísticos relacionados con cada modelo teórico, y las predicciones teóricas sobre los efectos primarios de las variables explicativas incluidas en cada modelo.

En el modelado de la severidad global se aplicaron técnicas de regresión robusta (Huber, 1981; Venables y Ripley, 2013). Los modelos de regresión robusta se ajustaron mediante un algoritmo de iteración de mínimos cuadrados ponderados (IWLS) utilizando la función `rlm` del paquete MASS de R. Las pruebas de significación global (todos los efectos nulos) y la prueba de significación parcial (efectos individuales nulos) se realizó mediante test de permutaciones basado en randomización (Edgington y Onghena, 2007) siguiendo la aproximación sugerida por Manly (1991) para modelos de regresión. Además, se presentaron también test de significación utilizados en regresión robusta

basados en teoría asintótica. Adicionalmente, los coeficientes de los modelos fueron estimados utilizando técnicas de “Bootstrapping pairs” e intervalos de confianza basados en bootstrap BCa (“Bias corrected and accelerated”). Este método para la estimación por intervalo de confianza de los coeficientes de regresión es poco sensible a desviaciones de supuestos del modelo estadístico, y permite obtener mejores estimaciones del error estándar de los parámetros estimados (Chernick, 2008; Efron y Tibshirani, 1994).



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación



Tabla 8.2: Modelos teóricos y predicciones teóricas sobre los efectos primarios de las variables explicativas en los modelos estadísticos de la severidad global. *CES* denota “cambio esperado en severidad”. *CES_{II,I}* denota el cambio en severidad al pasar del patrón atribucional II al I.

| Modelo Teórico | Variables explicativas en el modelo estadístico. | Fenómenos reflejados | Efectos primarios esperados en la severidad global de acuerdo al modelo. |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Modelo SDT (componente 1) | MOTIE (motivación intrínseca), VALUE (Valor/Utilidad, relacionada con nivel de integración y diferentes niveles de autonomía en la regulación de la motivación extrínseca). | Tipos de motivación y regulación de la motivación, y su relación con el efecto facilitador de la enseñanza y severidad de los estudiantes. Mejor aprendizaje y rendimiento; mayor efecto facilitador de la enseñanza. | Efectos primarios negativos de MOTIE y VALUE en la severidad global del estudiante. |
| Modelo TAS | AUTEE (autoeficacia para aprender), ERESE (expectativas positivas de resultados) | Diferencias en el aprendizaje y rendimiento; diferencias en el interés y percepción del valor de la enseñanza diferencias en el efecto facilitador de la enseñanza. | Efectos primarios negativos de AUTEE y ERESE en la severidad global del estudiante. |
| Modelo Atribucional | LOCUE (Locus de causalidad), ESTAE (Estabilidad) y CONTE (Controlabilidad) | Tipo de respuestas emocionales y que se reflejan en diferencias en el proceso de aprendizaje, y el efecto facilitador de la enseñanza. | Efectos primarios negativos de LOCUE, ESTAE y CONTE. $CES_{II,I} < 0$ |
| Modelo de Auto-teorías | AUTEE (autoeficacia para aprender), ELEME (elección de metas: preferencia por metas de aprendizaje respecto de metas de desempeño, relacionadas con la adopción de una auto-teoría de | Diferencias en el procesamiento de información, respuestas emocionales y comportamentales; diferencias en el efecto facilitador de la enseñanza. | Efectos primarios negativos de AUTEE y ELEME. |

8.3. Resultados.

8.3.1. Ajuste global entre datos y el modelo LLT-MFRM.

Los resultados reportados en esta sección corresponden al ajuste global obtenido entre los datos y el modelo LLT-MFRM. Cada sub-escala fue analizada por separado. En el análisis diagnóstico del modelo no fue necesario eliminar elementos de las facetas consideradas. Las proporciones de observaciones con residuos extremos (PRE_3 y PRE_2) resultó dentro de niveles aceptables, que no implican una degradación de la escala o en las medidas de los elementos.

Se encontró un ajuste global aceptable entre los datos y el modelo para todas las sub-escalas consideradas (Tabla 8.3). Los valores de *Infit* y *Outfit* para todas las escalas muestran un ajuste satisfactorio (Linacre, 2014). Además, los valores obtenidos de *Infit* y *Outfit* muestran también que la aleatoriedad observada en los patrones de respuesta son consistentes con el modelo. Los valores de *Outfit* son incluso mejores que los obtenidos para el modelo MFRM original, mientras que los valores de *Infit* sugieren un ajuste algo menor para el modelo LLT-MFRM. Globalmente, los valores *Infit* y *Outfit* muestran un nivel de ajuste similar, a pesar de que las restricciones lineales introducidas implican una disminución notable del número de elementos de medición considerados.

Adicionalmente, los valores encontrados de la media ($Media(Z_{Res})$) y desvío estándar ($SD(Z_{Res})$), de los residuos estandarizados (Z_{Res}) para el modelo, son muy próximos a los valores esperados para un ajuste perfecto entre los datos y el modelo ($Media(Z_{Res}) = 0$ y $SD(Z_{Res}) = 1$). Los valores de $Media(Z_{Res})$ y $SD(Z_{Res})$ son similares a los encontrados para el modelo MFRM original, si bien, indican un ajuste algo menor. Nuevamente, estos índices de ajuste muestran un ajuste similar entre los datos y modelos para el modelo LLT-MFRM y el modelo MFRM original.

Por su parte, cuando se analiza el ajuste absoluto global entre los datos y el modelo desde una perspectiva que prioriza la exactitud y precisión en la predicción de los puntajes observados, se encuentra en general un ajuste aceptable. No obstante, la diferencia en este caso entre el ajuste del modelo LLT-MFRM y el MFRM original es algo mayor. Si bien, el ajuste indicado por estas medidas diagnósticas es en general aceptable, la escala TRDF muestra

un valor de $CV(RMSE)$ que sugiere un ajuste pobre. En este caso la variabilidad no explicada por el modelo es un 25,43% de la media de los puntajes observados.

Los valores de PVE en la Tabla 8.3 describen la proporción de la varianza de los puntajes observados explicada por el modelo. Como se aprecia en la Tabla 8.3 los valores de PVE oscilan entorno a un 27% de la varianza de los puntajes observados. Las escalas con valores menores de PVE son TRDF (11%) y EVAL (22%), mientras que, las que tienen un mayor valor de PVE son RELP (34%) y VALR (32%). Por tanto, una proporción substancial de la varianza observada es explicada por el modelo. Los valores de PVE representan un tamaño de efecto entorno al medio, para las escalas TRDF y EVAL, mientras que para las demás, el tamaño de efecto se interpreta como elevado (J. Cohen, 1992; Linacre, 2005). Por tanto, si bien los valores de PVE son menores para el modelo LLT-MFRM que para el modelo original, estos siguen representando un tamaño de efecto elevado y explican una proporción elevada de la variabilidad observada.

En síntesis, el análisis del ajuste global del modelo LLT-MFRM indica que es aceptable. Adicionalmente, la disminución poco pronunciada en el ajuste global, respecto al modelo MFRM original, sugiere que las restricciones lineales en la severidad en el modelo LLT-MFRM capturan razonablemente el papel de las variables socio-cognitivas en el proceso de respuesta de los estudiantes. Por tanto, los resultados del ajuste global son consistentes con la hipótesis central de esta investigación.

Tabla 8.3: Ajuste global entre los datos y el modelo LLT-MFRM. Los valores normalizados ($CV(RMSE)$ y $NMBE$) son expresados como porcentaje respecto a la media de los puntajes observados.

| | VALR | INTR | RELP | TRDF | CLOR | ENTC | APPR | TRPR | EVAL |
|----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <i>Infit</i> | 0,96 | 0,97 | 0,97 | 1,02 | 0,97 | 0,98 | 0,97 | 0,96 | 1,00 |
| <i>Outfit</i> | 0,96 | 0,98 | 0,99 | 1,01 | 0,97 | 0,98 | 0,97 | 0,97 | 1,00 |
| <i>Media</i> (Z_{Res}) | -0,02 | -0,02 | -0,01 | 0,02 | -0,02 | -0,01 | -0,02 | -0,03 | -0,01 |
| <i>SD</i> (Z_{Res}) | 0,98 | 0,99 | 0,99 | 1,01 | 0,98 | 0,99 | 0,99 | 0,99 | 1,00 |
| PRE_3 | 0,01 | 0,01 | 0,01 | 0,00 | 0,01 | 0,01 | 0,01 | 0,01 | 0,00 |
| PRE_2 | 0,04 | 0,04 | 0,04 | 0,06 | 0,04 | 0,04 | 0,04 | 0,05 | 0,05 |
| PVE | 0,32 | 0,25 | 0,34 | 0,11 | 0,33 | 0,30 | 0,27 | 0,28 | 0,22 |
| MAE | 0,53 | 0,57 | 0,55 | 0,61 | 0,47 | 0,58 | 0,54 | 0,46 | 0,63 |
| RMSE | 0,68 | 0,72 | 0,74 | 0,81 | 0,68 | 0,73 | 0,69 | 0,67 | 0,80 |
| $CV(RMSE)$ | 18,13 | 19,44 | 17,96 | 25,43 | 17,96 | 19,37 | 18,45 | 18,36 | 19,66 |
| $NMBE$ | -0,43 | -0,36 | -0,25 | 0,51 | -0,32 | -0,23 | -0,34 | -0,49 | -0,21 |
| N° Obs. | 38265 | 42198 | 34910 | 27268 | 54736 | 34590 | 35900 | 27984 | 47145 |

8.3.2. Análisis de la Faceta Componentes de Severidad.

En las Tablas 8.4 y 8.5 se muestran, para cada una de las sub-escalas de calidad en la enseñanza, las medidas obtenidas para cada componente de severidad, sus errores estándar de medición y estadísticos de ajuste. Los valores de Infit y Outfit obtenidos para los componentes de severidad indican en general un buen nivel de ajuste de los elementos de esta faceta. Por otra parte, los valores de correlación punto medida (r_{pm}) mostraron consistencia entre las medidas de cada componente y el constructo medido por su sub-escala, para todos los componentes y sub-escalas consideradas. Observar que de 81 valores para los nueve componentes correspondientes a las nueve sub-escalas, 68 presentan una medida significativamente distinta de cero, como lo indican las medidas estimadas y los valores p de Holm obtenidos.

A pesar de este patrón general, las sub-escalas TRDF y EVAL presentan varios componentes (AUTEE[C1], ERESE[C2] y VALUE[C5]) con un ajuste menor que el observado para los mismos componentes, en las restantes sub-escalas. Esto podría sugerir diferencias en el proceso de respuesta de estas sub-escalas respecto de las otras. Por otra parte, los componentes LOCUE[C7] y ESTAE[C8] tienen un menor ajuste para la sub-escala TRDF que para todas las demás, lo que sugiere diferencias adicionales en el proceso de respuesta a esta sub-escala respecto a las restantes sub-escalas.

El análisis de la confiabilidad de las medidas de los componentes de severidad se presenta en la Tabla 8.6. El resultado obtenido en la prueba de efectos fijos (todos iguales) fue altamente significativo en todos los casos, lo que significa que al menos dos componentes tienen diferente severidad, para cada una de las sub-escalas consideradas. Como se aprecia en la Tabla 8.6 el valor del *RMSE* de la medida de los componentes es muy bajo y los índices de confiabilidad (*SR*, *G* y *H*) corresponden a muy altos niveles de confiabilidad de las medidas. En particular el número de estratos estadísticamente discernibles que podrían definirse para las medidas de severidad de los componentes varía aproximadamente entre 28 y 43, lo que indica una notable confiabilidad en la medición de los componentes.

Adicionalmente, en las Tablas 8.4 y 8.5 se puede apreciar que todos los componentes tienen una medida significativamente distinta de cero para varias de las sub-escalas de calidad en la enseñanza. En particular, los componentes AUTEE[C1], ERESE[C2], VALUE[C5] y CONTE[C9] contribuyen a las medidas

de severidad de los estudiantes para todas las sub-escalas consideradas. En forma similar, todas las sub-escalas tienen componentes de severidad con medidas significativamente distintas de cero. En especial, todos los componentes de severidad están implicados en la medida de severidad de los estudiantes correspondiente a la sub-escala VALR.

En síntesis, los resultados presentados en las Tablas 8.4, 8.5 y 8.6 son altamente consistentes con la hipótesis central de esta tesis. Esto es, las medidas de componentes de severidad, sus estadísticos de ajuste y confiabilidad, aportan evidencias de que las variables socio-cognitivas consideradas, que median cognitivamente el proceso de respuesta, o bien, están relacionadas con diferencias en el proceso de aprendizaje y el efecto facilitador de la enseñanza, explican diferencias en la severidad de los estudiantes al evaluar la enseñanza.



Tabla 8.4: Componentes de Severidad.

| Escala | Factores | Medida | S.E. | Infit | Outfit | r_{pm} | t | p | $PHolm$ |
|--------|-----------|--------|------|-------|--------|----------|--------|------|---------|
| VALR | AUTEE[C1] | -0,10 | 0,00 | 0,97 | 0,97 | 0,55 | -25,00 | 0,00 | 0,00 |
| | ERESE[C2] | -0,19 | 0,00 | 0,98 | 0,98 | 0,55 | -47,50 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,19 | 0,00 | 0,94 | 0,94 | 0,55 | -47,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,05 | 0,01 | 0,95 | 0,95 | 0,56 | 5,00 | 0,00 | 0,00 |
| | VALUE[C5] | -0,23 | 0,00 | 0,96 | 0,96 | 0,55 | -57,50 | 0,00 | 0,00 |
| | ELEME[C6] | 0,03 | 0,01 | 0,94 | 0,95 | 0,57 | 3,00 | 0,00 | 0,00 |
| | LOCUE[C7] | -0,04 | 0,01 | 0,96 | 0,96 | 0,56 | -4,00 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,16 | 0,01 | 0,94 | 0,95 | 0,56 | 16,00 | 0,00 | 0,00 |
| | CONTE[C9] | 0,21 | 0,01 | 0,97 | 0,97 | 0,56 | 21,00 | 0,00 | 0,00 |
| INTR | AUTEE[C1] | -0,04 | 0,00 | 1,01 | 1,01 | 0,48 | -10,00 | 0,00 | 0,00 |
| | ERESE[C2] | -0,19 | 0,00 | 1,01 | 1,02 | 0,47 | -47,50 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,15 | 0,00 | 0,97 | 0,98 | 0,49 | -37,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,02 | 0,00 | 0,97 | 0,97 | 0,50 | 5,00 | 0,00 | 0,00 |
| | VALUE[C5] | -0,12 | 0,00 | 0,99 | 1,00 | 0,48 | -30,00 | 0,00 | 0,00 |
| | ELEME[C6] | 0,01 | 0,00 | 0,96 | 0,97 | 0,51 | 2,50 | 0,01 | 0,13 |
| | LOCUE[C7] | -0,17 | 0,00 | 0,98 | 0,99 | 0,49 | -42,50 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,07 | 0,00 | 0,96 | 0,97 | 0,49 | 17,50 | 0,00 | 0,00 |
| | CONTE[C9] | 0,28 | 0,00 | 0,99 | 1,00 | 0,49 | 70,00 | 0,00 | 0,00 |
| RELP | AUTEE[C1] | -0,07 | 0,00 | 0,99 | 1,00 | 0,56 | -17,50 | 0,00 | 0,00 |
| | ERESE[C2] | -0,23 | 0,00 | 0,98 | 1,00 | 0,54 | -57,50 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,13 | 0,00 | 0,95 | 0,97 | 0,56 | -32,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,01 | 0,01 | 0,96 | 0,97 | 0,57 | 1,00 | 0,32 | 1,00 |
| | VALUE[C5] | -0,13 | 0,00 | 0,97 | 0,99 | 0,55 | -32,50 | 0,00 | 0,00 |
| | ELEME[C6] | -0,05 | 0,01 | 0,96 | 0,97 | 0,58 | -5,00 | 0,00 | 0,00 |
| | LOCUE[C7] | -0,14 | 0,01 | 0,97 | 0,98 | 0,57 | -14,00 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,02 | 0,01 | 0,95 | 0,96 | 0,56 | 2,00 | 0,05 | 0,00 |
| | CONTE[C9] | 0,20 | 0,01 | 0,98 | 0,99 | 0,57 | 20,00 | 0,00 | 0,00 |
| TRDF | AUTEE[C1] | 0,11 | 0,00 | 1,10 | 1,10 | 0,33 | 27,50 | 0,00 | 0,00 |
| | ERESE[C2] | -0,03 | 0,00 | 1,11 | 1,11 | 0,33 | -7,50 | 0,00 | 0,00 |
| | MOTIE[C3] | 0,01 | 0,00 | 1,06 | 1,06 | 0,34 | 2,50 | 0,01 | 0,13 |
| | PRELE[C4] | 0,04 | 0,01 | 1,05 | 1,05 | 0,33 | 4,00 | 0,00 | 0,00 |
| | VALUE[C5] | -0,04 | 0,00 | 1,08 | 1,08 | 0,34 | -10,00 | 0,00 | 0,00 |
| | ELEME[C6] | 0,11 | 0,01 | 1,03 | 1,03 | 0,34 | 11,00 | 0,00 | 0,00 |
| | LOCUE[C7] | 0,08 | 0,01 | 1,06 | 1,06 | 0,33 | 8,00 | 0,00 | 0,00 |
| | ESTAE[C8] | -0,35 | 0,01 | 1,02 | 1,02 | 0,33 | -35,00 | 0,00 | 0,00 |
| | CONTE[C9] | -0,17 | 0,01 | 1,06 | 1,06 | 0,33 | -17,00 | 0,00 | 0,00 |
| CLOR | AUTEE[C1] | -0,09 | 0,00 | 0,98 | 0,98 | 0,55 | -22,50 | 0,00 | 0,00 |
| | ERESE[C2] | -0,19 | 0,00 | 0,98 | 0,98 | 0,54 | -47,50 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,16 | 0,00 | 0,95 | 0,95 | 0,56 | -40,00 | 0,00 | 0,00 |
| | PRELE[C4] | 0,03 | 0,00 | 0,96 | 0,97 | 0,56 | 7,50 | 0,00 | 0,00 |
| | VALUE[C5] | -0,16 | 0,00 | 0,96 | 0,96 | 0,55 | -40,00 | 0,00 | 0,00 |
| | ELEME[C6] | 0,01 | 0,00 | 0,96 | 0,96 | 0,58 | 2,50 | 0,01 | 0,13 |
| | LOCUE[C7] | -0,12 | 0,00 | 0,98 | 0,97 | 0,56 | -30,00 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,05 | 0,00 | 0,95 | 0,95 | 0,56 | 12,50 | 0,00 | 0,00 |
| | CONTE[C9] | 0,20 | 0,00 | 0,98 | 0,98 | 0,56 | 50,00 | 0,00 | 0,00 |

Tabla 8.5: Componentes de Severidad (continuación).

| Escala | Factores | Medida | S.E. | Infit | Outfit | r_{pm} | t | p | $PHolm$ |
|--------|-----------|--------|------|-------|--------|----------|--------|------|---------|
| ENTC | AUTEE[C1] | -0,08 | 0,00 | 1,01 | 1,01 | 0,52 | -20,00 | 0,00 | 0,00 |
| | ERESE[C2] | -0,18 | 0,00 | 1,01 | 1,01 | 0,51 | -45,00 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,13 | 0,00 | 0,98 | 0,98 | 0,52 | -32,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,03 | 0,01 | 0,98 | 0,99 | 0,53 | 3,00 | 0,00 | 0,00 |
| | VALUE[C5] | -0,11 | 0,00 | 1,00 | 1,00 | 0,52 | -27,50 | 0,00 | 0,00 |
| | ELEME[C6] | -0,01 | 0,01 | 0,98 | 0,99 | 0,55 | -1,00 | 0,32 | 1,00 |
| | LOCUE[C7] | -0,17 | 0,01 | 1,00 | 1,00 | 0,53 | -17,00 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,00 | 0,01 | 0,97 | 0,97 | 0,52 | 0,00 | 1,00 | 1,00 |
| | CONTE[C9] | 0,24 | 0,01 | 1,00 | 1,00 | 0,53 | 24,00 | 0,00 | 0,00 |
| APPR | AUTEE[C1] | -0,07 | 0,00 | 1,00 | 1,00 | 0,51 | -17,50 | 0,00 | 0,00 |
| | ERESE[C2] | -0,20 | 0,00 | 0,99 | 1,00 | 0,50 | -50,00 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,15 | 0,00 | 0,97 | 0,97 | 0,51 | -37,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,01 | 0,01 | 0,97 | 0,97 | 0,51 | 1,00 | 0,32 | 1,00 |
| | VALUE[C5] | -0,13 | 0,00 | 0,98 | 0,98 | 0,50 | -32,50 | 0,00 | 0,00 |
| | ELEME[C6] | 0,01 | 0,01 | 0,96 | 0,97 | 0,53 | 1,00 | 0,32 | 1,00 |
| | LOCUE[C7] | -0,11 | 0,01 | 0,98 | 0,99 | 0,51 | -11,00 | 0,00 | 0,00 |
| | ESTAE[C8] | 0,05 | 0,01 | 0,95 | 0,96 | 0,51 | 5,00 | 0,00 | 0,00 |
| | CONTE[C9] | 0,24 | 0,01 | 0,99 | 0,99 | 0,51 | 24,00 | 0,00 | 0,00 |
| TRPR | AUTEE[C1] | -0,10 | 0,01 | 1,00 | 1,00 | 0,51 | -10,00 | 0,00 | 0,00 |
| | ERESE[C2] | -0,20 | 0,01 | 1,00 | 1,00 | 0,51 | -20,00 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,22 | 0,00 | 0,96 | 0,97 | 0,52 | -55,00 | 0,00 | 0,00 |
| | PRELE[C4] | -0,01 | 0,01 | 0,97 | 0,97 | 0,53 | -1,00 | 0,32 | 1,00 |
| | VALUE[C5] | -0,16 | 0,01 | 1,01 | 1,01 | 0,50 | -16,00 | 0,00 | 0,00 |
| | ELEME[C6] | 0,02 | 0,01 | 0,97 | 0,98 | 0,54 | 2,00 | 0,05 | 0,40 |
| | LOCUE[C7] | -0,01 | 0,01 | 0,97 | 0,98 | 0,52 | -1,00 | 0,32 | 1,00 |
| | ESTAE[C8] | 0,20 | 0,01 | 0,96 | 0,96 | 0,52 | 20,00 | 0,00 | 0,00 |
| | CONTE[C9] | 0,21 | 0,01 | 0,98 | 0,98 | 0,52 | 21,00 | 0,00 | 0,00 |
| EVAL | AUTEE[C1] | -0,06 | 0,00 | 1,06 | 1,06 | 0,48 | -15,00 | 0,00 | 0,00 |
| | ERESE[C2] | -0,05 | 0,00 | 1,07 | 1,06 | 0,49 | -12,50 | 0,00 | 0,00 |
| | MOTIE[C3] | -0,19 | 0,00 | 1,03 | 1,03 | 0,47 | -47,50 | 0,00 | 0,00 |
| | PRELE[C4] | 0,02 | 0,00 | 1,02 | 1,02 | 0,48 | 5,00 | 0,00 | 0,00 |
| | VALUE[C5] | -0,03 | 0,00 | 1,07 | 1,06 | 0,49 | -7,50 | 0,00 | 0,00 |
| | ELEME[C6] | 0,04 | 0,00 | 1,01 | 1,01 | 0,49 | 10,00 | 0,00 | 0,00 |
| | LOCUE[C7] | 0,01 | 0,00 | 1,03 | 1,03 | 0,48 | 2,50 | 0,01 | 0,13 |
| | ESTAE[C8] | 0,01 | 0,00 | 1,01 | 1,01 | 0,47 | 2,50 | 0,01 | 0,13 |
| | CONTE[C9] | 0,14 | 0,00 | 1,02 | 1,02 | 0,48 | 35,00 | 0,00 | 0,00 |

Tabla 8.6: Confiabilidad de Separación para la faceta Componentes de Severidad para cada sub-escala.

| | $RMSE$ | SR | G | H | Prueba Efectos Fijos (todos iguales) valor p |
|----------|--------|------|-------|-------|------------------------------------------------|
| VALR(E1) | 0,00 | 1,00 | 30,57 | 41,10 | 0,00 |
| INTR(E2) | 0,00 | 1,00 | 32,24 | 43,32 | 0,00 |
| RELP(E3) | 0,00 | 1,00 | 23,38 | 31,50 | 0,00 |
| TRDF(E4) | 0,00 | 1,00 | 38,51 | 28,63 | 0,00 |
| CLOR(E5) | 0,00 | 1,00 | 29,56 | 39,75 | 0,00 |
| ENTC(E6) | 0,00 | 1,00 | 25,76 | 34,68 | 0,00 |
| APPR(E7) | 0,00 | 1,00 | 25,28 | 34,04 | 0,00 |
| TRPR(E8) | 0,01 | 1,00 | 25,92 | 34,90 | 0,00 |
| EVAL(E9) | 0,00 | 1,00 | 22,07 | 29,76 | 0,00 |

Por otra parte, el patrón de valores de las medidas de los componentes, para cada sub-escala, puede informarnos adicionalmente del papel que cumplen los componentes de severidad, y variables socio-cognitivas relacionadas, en el proceso de respuesta de los estudiantes al evaluar la enseñanza. Las medidas de los componentes son resumidos por sub-escala en la Tabla 8.7, para mostrar la distribución de las medidas. La severidad promedio de los componentes por sub-escala es negativa (entre -0,01 y -0,06 logits). Por ejemplo, para VALR, que tiene una severidad media de los componentes igual a -0,03 logits tiene un rango de severidades estimado para los estudiantes de -2,32 a -0,3. El patrón de las medidas de los componentes de severidad para las diferentes sub-escalas se representa en la Figura 8.1.

Tabla 8.7: Resumen de las medidas de los Componentes de Severidad.

| | VALR | INTR | RELP | TRDF | CLOR | ENTC | APPR | TRPR | EVAL |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Media | -0,03 | -0,03 | -0,06 | -0,03 | -0,05 | -0,05 | -0,04 | -0,03 | -0,01 |
| SD | 0,15 | 0,14 | 0,12 | 0,14 | 0,12 | 0,12 | 0,13 | 0,15 | 0,08 |
| Q1 | -0,19 | -0,15 | -0,13 | -0,04 | -0,16 | -0,13 | -0,13 | -0,16 | -0,05 |
| Md | -0,04 | -0,04 | -0,07 | 0,01 | -0,09 | -0,08 | -0,07 | -0,01 | 0,01 |
| Q3 | 0,05 | 0,02 | 0,01 | 0,08 | 0,03 | 0 | 0,01 | 0,02 | 0,02 |

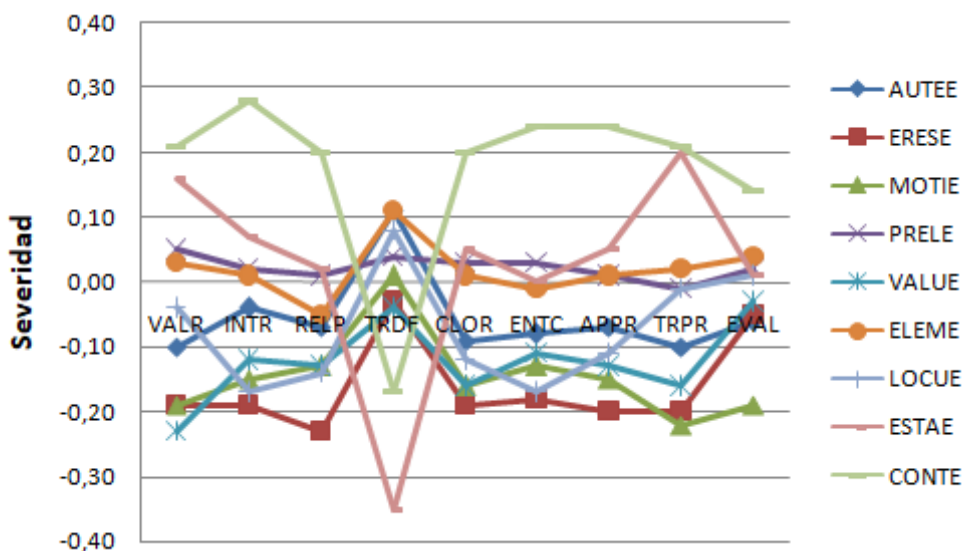


Figura 8.1: Medida de los componentes de severidad para cada sub-escala de evaluación de la calidad en la enseñanza.

Los patrones mostrados en la Figura 8.1 se presentan organizados en la Figura 8.2 de acuerdo a los modelos teóricos propuestos en el capítulo 4 (véase Tabla 8.2). Para interpretar correctamente los patrones que se muestran en la Figura 8.2 es importante entender que los modelos teóricos considerados describen el proceso de respuesta desde cierta perspectiva teórica, y considerando simultáneamente sólo algunas de las variables vinculadas a los componentes o factores de severidad. Así, los efectos primarios mencionados en la Tabla 8.2 corresponden al efecto de cierta variable socio-cognitiva, corregida por las demás variables consideradas en el modelo.

Por el contrario, las medidas de los componentes obtenidas mediante el modelo LLT-MFRM se relacionan con los efectos primarios de cada variable socio-cognitiva, corregida por todas las demás consideradas (incluyendo algunas que no están en el modelo teórico). Adicionalmente, en su conjunto las variables socio-cognitivas consideradas tienen un alto nivel de interrelación e interdependencia. Esto implica, que los efectos primarios referidos en los modelos teóricos, pueden o no corresponderse con las medidas de los componentes de severidad obtenidos con el modelo LLT-MFRM.

No obstante la aclaración realizada previamente, resultó útil analizar los patrones de medidas de los factores de severidad, desde la perspectiva de los modelos teóricos y compararlos con los efectos primarios esperados de las variables socio-cognitivas.

En la Figura 8.2 se puede apreciar un acuerdo casi perfecto entre los patrones observados y los esperados de acuerdo a los modelos SDT y TAS. Cierta discrepancia parece existir en el patrón observado para la sub-escala TRDF. Esto parece sugerir que el proceso de respuesta a esta escala podría diferir respecto al proceso de respuesta de las otras sub-escalas. Lo anterior es consistente con las diferencias de ajuste que se señalaron para varios componentes de severidad entre esta sub-escala y las demás.

Por otra parte, los patrones esperados de acuerdo al modelo Atribucional parecen diferir de los patrones de medidas de los componentes de severidad. Si bien, se observa un acuerdo claro en el patrón observado para el componente LOCUE[C7], con el modelo atribucional, los otros dos componentes vinculados a este modelo (ESTAE[C8] y CONTE[C9]) presentan un patrón diferente del esperado. Esta diferencia, podría ser debida a que los efectos de las variables ESTAE y CONTE, cuando se ajustan por todas las demás, difieren del efecto primario de estas variables cuando solo se ajustan por las variables del modelo

Atribucional.

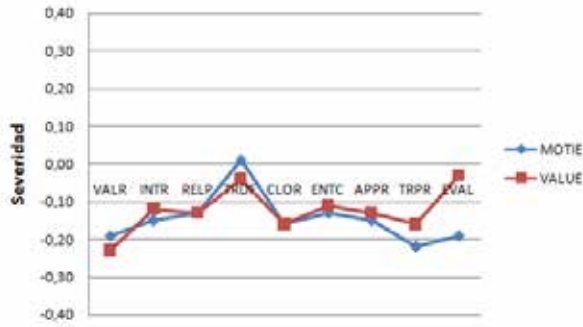
Otra alternativa, podría ser que en este caso las diferencias de severidad vinculadas a estas variables no estén de acuerdo con un patrón de validez en el proceso de respuesta de los estudiantes. Esto es, las variables ESTAE y CONTE podrían explicar diferencias en severidad, opuestas en dirección al efecto de estas variables en el proceso de aprendizaje y del efecto facilitador de la enseñanza. Es decir, estudiantes con una mejor calidad de aprendizaje, para los que la enseñanza tienen un efecto facilitador mayor, no obstante, serían más severos al evaluar la enseñanza.

Un vez más, la sub-escala TRDF presenta un patrón diferente en las medidas de los componentes de severidad que sugiere diferencias en el patrón de respuesta a dicha sub-escala. Observar que el patrón para las medidas de los componentes de esta sub-escala (TRDF: LOCUE +, ESTAE -, CONTE -) es el opuesto al observado en las otras sub-escalas (LOCUE -, ESTAE +, CONTE+).

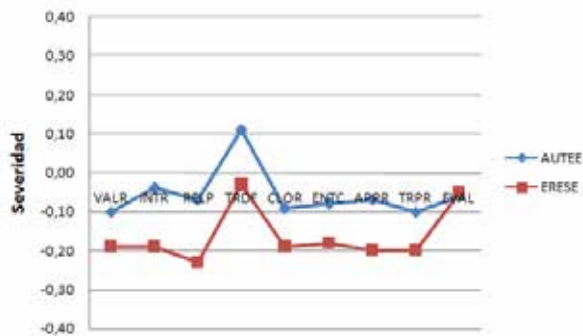
Por último, en la Figura 8.2 se presentan los patrones vinculados al modelo de Auto-teorías. Se observa un acuerdo perfecto entre el patrón esperado y el observado para el componente AUTE(C1) y el patrón opuesto para ELEME[C6]. Explicaciones similares a las dadas para los componentes ESTAE[8] y CONTE[9] en el modelo Atribucional, podrían aplicar en este caso a la discrepancia observada entre el patrón esperado a partir del modelo de Auto-teorías y el patrón observado.

En la Figura 8.3 se representan los patrones de respuesta, particionando las sub-escalas y componentes en dos sub-grupos. El primer grupo de sub-escalas (Grupo E1) incluye las sub-escalas VALR, INTR, RELP, ENTC, APPR Y TRPR, mientras que, el segundo grupo de sub-escalas (Grupo E2) está formado por las sub-escalas TRDF y EVAL. Este agrupamiento parece razonable debido a las evidencias que se han encontrado de posibles diferencias en el proceso de respuesta, especialmente de la sub-escala TRDF respecto a las demás sub-escalas.

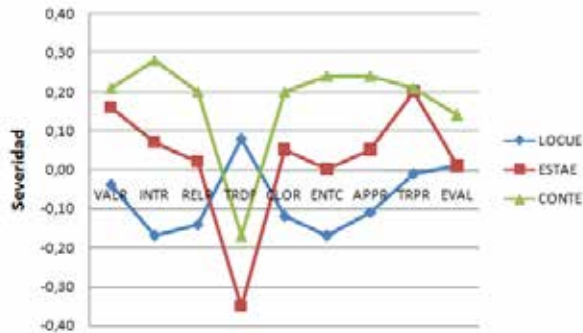
Por otra parte, los componentes de severidad se particionaron en los grupos: AUTE(C1), ERESE[C2], MOTIE[C3], VALUE[C5] y LOCUE[C7] (Grupo C1); y, PRELE[C4], ELEME[C6], ESTAE[C8] y CONTE[C9] (Grupo C2). Estos dos grupos de componentes de severidad parecen también diferir entre sí, en especial, respecto a su relación con la severidad de los estudiantes.



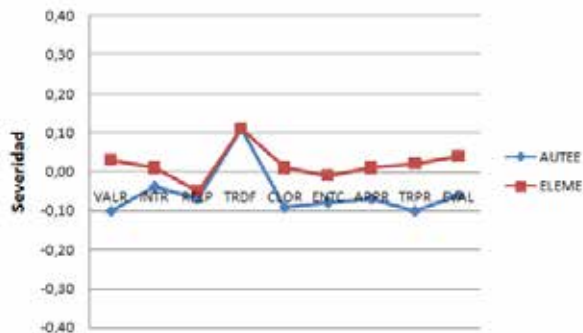
(a) Modelo SDT.



(b) Modelo TAS.



(c) Modelo Atribucional.



(d) Modelo de Auto-teorías.

Figura 8.2: Modelos teóricos del proceso de respuesta a los SET y componentes de severidad para las sub-escalas de calidad en la enseñanza.

En la Figura 8.3a se muestran los patrones observados para el grupo E1 y grupo C1, que están en perfecto acuerdo con los patrones esperados de acuerdo a los modelos teóricos. Por otra parte, los patrones observados para el grupo E2 y grupo C1, que se presentan en la Figura 8.3b, muestran un acuerdo para la sub-escala EVAL entre el patrón observado y el esperado de acuerdo a los modelos teóricos. Por el contrario, para la sub-escala TRDF, el patrón es opuesto al observado para la variable EVAL, y se encuentra en desacuerdo con lo esperado de acuerdo a los modelos teóricos. Esta observación, una vez más, señala diferencias en el proceso de respuesta de la sub-escala TRDF respecto a las demás.

Adicionalmente, en la Figura 8.3c se aprecia que el grupo E1 de sub-escalas tienen patrones bastante similares, también, respecto al grupo C2 de componentes. No obstante la Figura 8.3c sugiere que las sub-escalas VALR y TRPR se diferencian de las demás sub-escalas con respecto al componente ESTAE(C8). Por último, los patrones para las sub-escalas TRDF y EVAL respecto al grupo C2 es similar para los componentes PRELE[C4] y ELEM[C6], pero, difieren bastante respecto de los componentes ESTAE[C8] Y CONTE[C9].

La discusión de los resultados presentados en la Figuras 8.2 y 8.3 muestra con claridad que existe una relación fundamental entre las variables socio-cognitivas consideradas y el proceso de respuesta a los cuestionarios de evaluación estudiantil de la calidad en la enseñanza. Esto constituye una evidencia adicional de la hipótesis central de esta investigación.

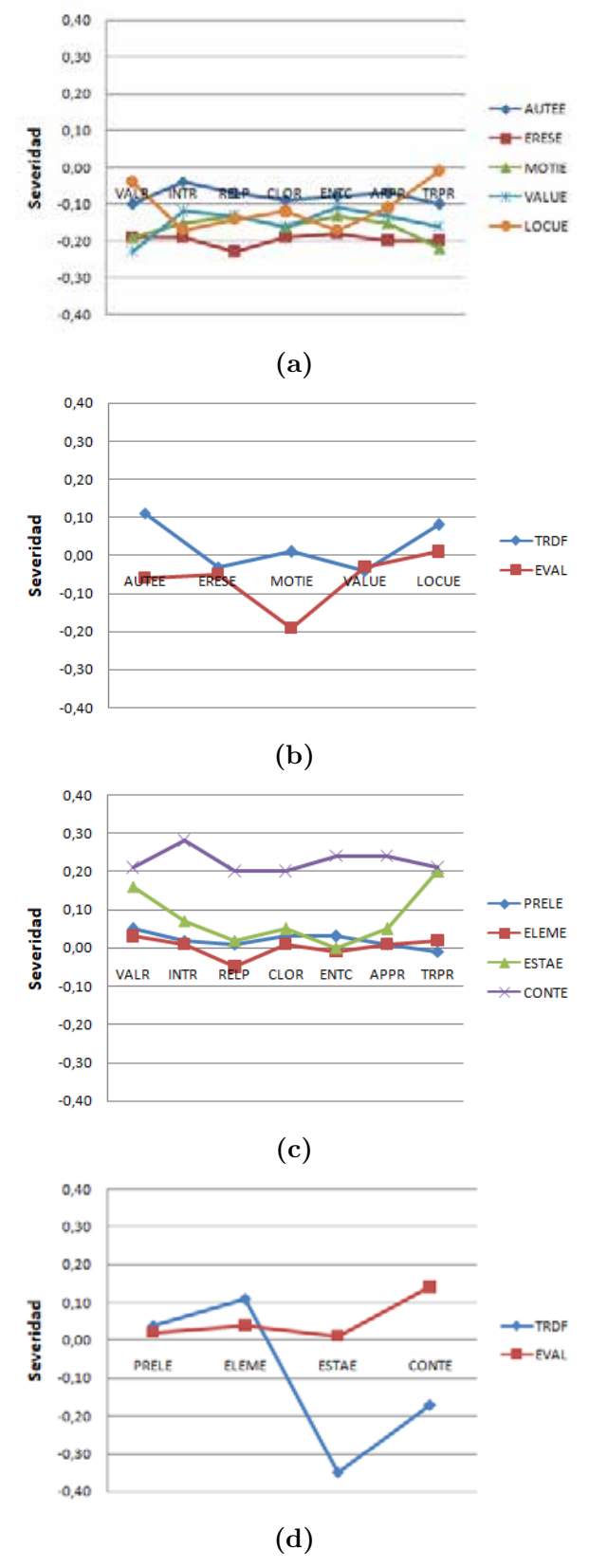


Figura 8.3: Agrupamiento de patrones.



8.3.3. Evaluación de la hipótesis de investigación: análisis correlacional.

De las 81 correlaciones consideradas, entre las medidas de las severidades de los estudiantes y las variables socio-cognitivas, 73 resultaron significativamente diferentes de cero como muestran los valores p de Holm corregidos para mantener un error de familia menor a 0,05 (véase Tabla 8.9). Esto es altamente consistente con el patrón predicho a partir de la hipótesis de investigación propuesta.

Por otra parte, de las 81 correlaciones consideradas 69 son significativamente menores que cero, como muestran los valores de las correlaciones y sus correspondientes valores p de Holm. Las medidas de severidad para cinco de las nueve sub-escalas de calidad en la enseñanza (VALRE, INTRE, RELPE, CLORE y TRPRE) tienen todas sus correlaciones negativas con las medidas socio-cognitivas consideradas.

Solo dos de las medidas de severidad tienen correlaciones positivas estadísticamente significativas (EVALE con 3/9 y TRDFE con 1/9). Por lo tanto, el patrón de correlaciones encontrado es altamente consistente con lo esperado de acuerdo a los modelos teóricos propuestos, si las sub-escalas consideradas son validas y las diferencias en severidad de los estudiantes reflejan adecuadamente diferencias en su proceso de aprendizaje y el efecto facilitador de la enseñanza. En otros términos, estudiantes que son menos severos al evaluar la enseñanza, a su vez estarían aprendiendo mejor y el efecto facilitador de la enseñanza sobre su aprendizaje sería mayor.

Tabla 8.8: Correlaciones (corregidas para atenuación) entre medidas de severidad y medidas de variables socio-cognitivas para cada sub-escala

| | AUTEE | ERESE | MOTIE | PRELE | VALUE | ELEME | LOCUE | ESTAE | CONTE |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VALRE | -0,51 | -0,56 | -0,62 | -0,38 | -0,60 | -0,10 | -0,21 | -0,21 | -0,13 |
| INTRE | -0,34 | -0,40 | -0,53 | -0,30 | -0,44 | -0,08 | -0,10 | -0,22 | -0,02 |
| RELPE | -0,35 | -0,40 | -0,39 | -0,26 | -0,36 | -0,09 | -0,18 | -0,16 | -0,12 |
| TRDFE | -0,22 | -0,08 | -0,11 | -0,21 | -0,05 | -0,14 | -0,26 | 0,21 | -0,27 |
| CLORE | -0,38 | -0,42 | -0,49 | -0,27 | -0,43 | -0,08 | -0,17 | -0,18 | -0,10 |
| ENTCE | -0,27 | -0,32 | -0,42 | -0,19 | -0,33 | -0,03 | -0,04 | -0,24 | 0,03 |
| APPRE | -0,34 | -0,41 | -0,52 | -0,28 | -0,46 | -0,06 | -0,05 | -0,29 | 0,04 |
| TRPRE | -0,50 | -0,54 | -0,55 | -0,39 | -0,53 | -0,11 | -0,25 | -0,19 | -0,18 |
| EVALE | -0,28 | -0,28 | -0,45 | -0,19 | -0,33 | 0,08 | 0,03 | -0,26 | 0,10 |

Tabla 8.9: Valor p (valor p de Holm) para las correlaciones (corregidas para atenuación) entre medidas de severidad y medidas de variables socio-cognitivas para cada sub-escala

| | AUTEE | ERESE | MOTIE | PRELE | VALUE | ELEME | LOCUE | ESTAE | CONTE |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| VALRE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) |
| INTRE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,39 (0,60) |
| RELPE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) |
| TRDFE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,02 (0,16) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) |
| CLORE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) |
| ENTCE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,15 (0,60) | 0,06 (0,36) | 0,00 (0,00) | 0,22 (0,60) |
| APPRE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,01 (0,09) | 0,03 (0,21) | 0,00 (0,00) | 0,07 (0,36) |
| TRPRE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) |
| EVALE | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,00 (0,00) | 0,20 (0,60) | 0,00 (0,00) | 0,00 (0,00) |

8.3.4. Evaluación de la hipótesis de investigación: modelado de la severidad global.

Ajuste global de los modelos.

El primer punto considerado en esta sección refiere al ajuste global de los modelos para cada una de las sub-escalas. La información aportada en relación a este punto es relevante para la evaluación de la hipótesis central de esta investigación y de los modelos teóricos propuestos en el capítulo 4. Adicionalmente, esta información puede ayudarnos en una exploración más profunda del proceso de respuesta de los estudiantes al evaluar la enseñanza.

Diferencias en el ajuste relacionadas con las sub-escalas, podrían reflejar diferencias vinculadas a las dimensiones de la calidad en la enseñanza medidas. A su vez, la evaluación de diferentes dimensiones de la enseñanza, probablemente, involucran distintas representaciones mentales que permiten procesar la información en las señales utilizadas para evaluar la enseñanza. Adicionalmente, diferentes dimensiones de la calidad en la enseñanza pueden relacionarse en forma distinta con el efecto facilitador de la enseñanza en el proceso de aprendizaje de los estudiantes.

Por otra parte, las diferencias de ajuste entre los modelos de regresión, podrían reflejar disparidades en los procesos que describen los modelos teóricos con los que están relacionados.

Así, el modelo SDT está primariamente relacionado con diferencias en los tipos de motivación y tipos de regulación de la motivación. Por su parte, el modelo TAS refleja sobre todo el efecto de diferencias en las expectativas de los estudiantes en relación a su autoeficacia para aprender y sus expectativas respecto a resultados positivos en su aprendizaje y rendimiento. A su vez, el modelo Atribucional se vincula con diferencias en las dimensiones de causalidad relacionadas con las causas atribuidas a éxitos o fracasos en el contexto de la enseñanza de un docente impartida en un curso. Por último, el modelo de Auto-teorías está relacionado con el papel de auto-teorías (crecimiento o entidad) en estudiantes con diferentes niveles de auto-eficacia, y su relación con la adopción de metas de aprendizaje o de desempeño.

A continuación se describen los resultados de ajuste de cada modelo para las escalas, así como también, las diferencias en el ajuste entre modelos y entre escalas.

En la Tabla 8.10 se muestran los índices de ajuste y los resultados del

test F global para los modelos de regresión correspondientes a cada modelo teórico considerado. Para los cuatro modelos y nueve sub-escalas consideradas se encontró una diferencia estadísticamente significativa. Por tanto, al menos uno de los efectos presentes en cada modelo es diferente de cero.

Unos de los aspectos analizados fue el ajuste absoluto de los modelos para cada sub-escala. Un buen ajuste absoluto del modelo a la escala se relaciona con una mayor precisión en la predicción de la severidad global. Los valores obtenidos de MAE y RMSE (y sus valores normalizados) muestran que el ajuste absoluto de los cuatro modelos, para cada una de las nueve sub-escalas, presenta por lo menos un ajuste aceptable. Esto significa que los modelos tienen una capacidad al menos aceptable de predicción de la severidad global.

No obstante, si bien el ajuste absoluto de los modelos para cada una de las sub-escalas es aceptable en general, para algunas escalas en particular el ajuste es mejor (véase Figura 8.4). Por ejemplo, los modelos SDT y TAS, presentan un ajuste bueno para las escalas RELP, CLOR, ENTC y EVAL, que es un poco mayor al ajuste encontrado para las otras escalas. Por otra parte, para el modelo atribucional, el ajuste mayor se presentó para las escalas RELP y ENTC. Por último, para el modelo de Auto-teorías, el ajuste es algo mayor para las escalas RELP, CLOR y ENTC. Por tanto, parecen existir diferencias en la capacidad de predicción de los modelos entre las dimensiones de calidad en la enseñanza medidas por las sub-escalas.

En la Tabla 8.10 se muestran también los valores de *MBE* y *NMBE* obtenidos para los modelos. Los valores de *MBE*, y su valor normalizado *NMBE*, describen la diferencia entre el promedio de los valores observados de severidad y el promedio de los valores predichos por los modelos. Cuando se encuentran valores grandes de *MBE* y *NMBE*, esto significa que las predicciones son sesgadas.

Los valores muy bajos de *MBE* encontrados indican un buen ajuste de los modelos, y desde un punto de vista práctico, la casi completa falta de sesgo en las predicciones. Valores de *NMBE* menores a 15% se consideran aceptables (Reddy, 2011; Reddy y cols., 1997). Los valores observados para los cuatro modelos, para las nueve sub-escalas consideradas, son en todos los casos menores a 1% (véase Figura 8.4e y 8.4f).

Tabla 8.10: Ajuste global entre los datos y modelos regresión correspondientes a cada modelo teórico. La variable respuesta es la severidad global. La selección de las variables explicativas se fundamenta en los modelos teóricos propuestos en el capítulo 4. Los valores normalizados ($NMAE$, $NRMSE$ y $NMBE$) son expresados como porcentaje respecto al rango de la severidad para cada dimensión. F_{Obs} es el valor observado del estadístico para un test F global. p_{Perm} es el valor p de permutación correspondiente al F_{Obs} .

| | | VALR | INTR | RELP | TRDF | CLOR | ENTC | APPR | TRPR | EVAL |
|---------------------------|---------------|--------|--------|--------|-------|--------|--------|--------|--------|-------|
| Modelo SDT | MAE | 1,41 | 1,45 | 2,07 | 1,62 | 1,80 | 1,64 | 1,63 | 1,99 | 1,24 |
| | $RMSE$ | 1,89 | 1,93 | 3,10 | 2,16 | 2,48 | 2,48 | 2,15 | 2,62 | 1,74 |
| | MBE | -0,02 | -0,04 | 0,12 | 0,06 | 0,03 | 0,05 | -0,05 | 0,03 | -0,08 |
| | $NMAE$ | 7,66 | 8,00 | 6,23 | 9,09 | 6,71 | 5,87 | 8,34 | 9,41 | 7,66 |
| | $NRMSE$ | 10,23 | 10,63 | 9,34 | 12,14 | 9,25 | 8,86 | 11,01 | 12,38 | 9,99 |
| | $NMBE$ | -0,11 | -0,20 | 0,36 | 0,35 | 0,11 | 0,17 | -0,23 | 0,12 | -0,44 |
| | R^2 | 0,38 | 0,27 | 0,15 | 0,01 | 0,23 | 0,15 | 0,25 | 0,31 | 0,08 |
| | R^2_{Ajust} | 0,38 | 0,27 | 0,15 | 0,01 | 0,23 | 0,15 | 0,25 | 0,31 | 0,08 |
| | F_{Obs} | 386,06 | 227,90 | 110,63 | 5,37 | 183,39 | 109,64 | 208,08 | 75,14 | 51,36 |
| | p_{Perm} | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Modelo TAS | MAE | 1,53 | 1,56 | 2,07 | 1,59 | 1,87 | 1,75 | 1,75 | 2,01 | 1,24 |
| | $RMSE$ | 2,04 | 2,08 | 3,11 | 2,12 | 2,58 | 2,57 | 2,30 | 2,69 | 1,74 |
| | MBE | 0,01 | -0,02 | 0,14 | 0,03 | 0,04 | 0,06 | -0,03 | 0,03 | -0,08 |
| | $NMAE$ | 8,30 | 8,58 | 6,24 | 8,96 | 6,98 | 6,25 | 8,97 | 9,52 | 7,11 |
| | $NRMSE$ | 11,05 | 11,43 | 9,35 | 11,90 | 9,61 | 9,18 | 11,78 | 12,70 | 9,99 |
| | $NMBE$ | 0,04 | -0,09 | 0,41 | 0,19 | 0,15 | 0,22 | -0,13 | 0,14 | -0,44 |
| | R^2 | 0,28 | 0,15 | 0,15 | 0,05 | 0,17 | 0,09 | 0,14 | 0,27 | 0,08 |
| | R^2_{Ajust} | 0,28 | 0,15 | 0,15 | 0,05 | 0,17 | 0,09 | 0,14 | 0,27 | 0,08 |
| | F_{Obs} | 242,90 | 112,68 | 107,80 | 30,19 | 124,04 | 60,16 | 103,68 | 230,48 | 51,36 |
| | p_{Perm} | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Modelo Atribucional | MAE | 1,76 | 1,64 | 2,26 | 1,58 | 2,00 | 1,81 | 1,82 | 2,29 | 1,25 |
| | $RMSE$ | 2,29 | 2,18 | 3,26 | 2,11 | 2,72 | 2,61 | 2,38 | 2,98 | 1,76 |
| | MBE | 0,01 | -0,03 | 0,14 | 0,06 | 0,03 | 0,06 | -0,03 | -0,03 | -0,08 |
| | $NMAE$ | 9,53 | 9,05 | 6,78 | 8,87 | 7,46 | 6,47 | 9,31 | 10,82 | 7,18 |
| | $NRMSE$ | 12,39 | 12,01 | 9,82 | 11,89 | 10,14 | 9,34 | 12,19 | 14,10 | 10,12 |
| | $NMBE$ | 0,05 | -0,18 | 0,41 | 0,34 | 0,11 | 0,20 | -0,14 | -0,12 | -0,45 |
| | R^2 | 0,10 | 0,07 | 0,06 | 0,05 | 0,07 | 0,06 | 0,08 | 0,10 | 0,05 |
| | R^2_{Ajust} | 0,10 | 0,07 | 0,06 | 0,05 | 0,07 | 0,06 | 0,08 | 0,10 | 0,05 |
| | F_{Obs} | 33,67 | 22,52 | 20,56 | 16,30 | 24,48 | 19,17 | 28,33 | 35,40 | 18,00 |
| | p_{Perm} | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Modelo de Auto-teorías | MAE | 1,63 | 1,62 | 2,17 | 1,58 | 1,95 | 1,80 | 1,80 | 2,13 | 1,25 |
| | $RMSE$ | 2,13 | 2,13 | 3,20 | 2,10 | 2,64 | 2,61 | 2,36 | 2,79 | 1,74 |
| | MBE | 0,02 | -0,02 | 0,15 | 0,04 | 0,05 | 7,08 | -0,02 | 0,02 | -0,08 |
| | $NMAE$ | 8,84 | 8,90 | 6,54 | 8,90 | 7,26 | 6,42 | 9,24 | 10,08 | 7,16 |
| | $NRMSE$ | 11,53 | 11,75 | 9,61 | 11,82 | 9,86 | 9,33 | 12,09 | 13,21 | 10,01 |
| | $NMBE$ | 0,09 | -0,10 | 0,46 | 0,25 | 0,18 | 0,25 | -0,13 | 0,07 | -0,44 |
| | R^2 | 0,22 | 0,11 | 0,10 | 0,06 | 0,12 | 0,06 | 0,10 | 0,21 | 0,07 |
| | R^2_{Ajust} | 0,22 | 0,11 | 0,10 | 0,06 | 0,12 | 0,06 | 0,10 | 0,21 | 0,07 |
| | F_{Obs} | 171,58 | 73,93 | 68,95 | 39,77 | 86,86 | 38,50 | 66,19 | 166,08 | 49,39 |
| | p_{Perm} | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |

En la Tabla 8.10 se presentan los valores obtenidos de R^2 y R^2_{Ajust} , que describen el ajuste relativo de los modelos. El valor de R^2 corresponde a la proporción de la varianza de la variable respuesta (severidad global) explicada por el modelo, mientras que R^2_{Ajust} corresponde al valor ajustado, tomando en cuenta el número de parámetros incluidos en el modelo. Los valores obtenidos de R^2 y R^2_{Ajust} muestran que existen diferencias en la proporción de la varianza

explicada entre modelos y entre sub-escalas.

Si se aplica la orientación general dada por [J. Cohen \(1992\)](#) para interpretar el tamaño del efecto representado por los valores de R^2 y R_{Ajus}^2 , es decir, 2% (tamaño del efecto pequeño), 13% (tamaño de efecto medio) y 26% (tamaño de efecto grande), se puede apreciar cierto orden entre los modelos de acuerdo a los tamaños de los efectos encontrados.

Si se ordenan los modelos para reflejar un orden decreciente de los tamaños de los efectos encontrados, el orden resultante sería: (1) modelo SDT; (2) modelo TAS; (3) modelo de Auto-teorías, y (4) modelo Atribucional. Este patrón se observa claramente en la [Figura 8.4g](#). Por ejemplo, si se consideran los tamaños de efecto elevado resulta: modelo SDT (3/9); modelo TAS (2/9); modelo de Auto-teorías (0/9); y, modelo Atribucional (0/9). Si se consideran los tamaños de efecto medios resulta: modelo SDT (4/9); modelo TAS (3/9); modelo de Auto-teorías (2/9); y, modelo Atribucional (0/9).

Los valores de R^2 y R_{Ajus}^2 difieren también entre sub-escalas como se aprecia en la [Tabla 8.10](#) y [Figura 8.4](#). En la [Figura 8.4](#) se aprecia que las sub-escalas VALR y TRPR presentan valores de R_{Ajus}^2 relativamente similares, para los cuatro modelos, y superiores a las demás sub-escalas. Por otra parte, las escalas TRDF y EVAL presentan también valores similares, pero, inferiores a los encontrados en las otras sub-escalas. Las demás sub-escalas presentan valores intermedios, similares para los modelos Atribucional y de Auto-teorías, mientras que con mayor variabilidad para los modelos SDT y TAS.

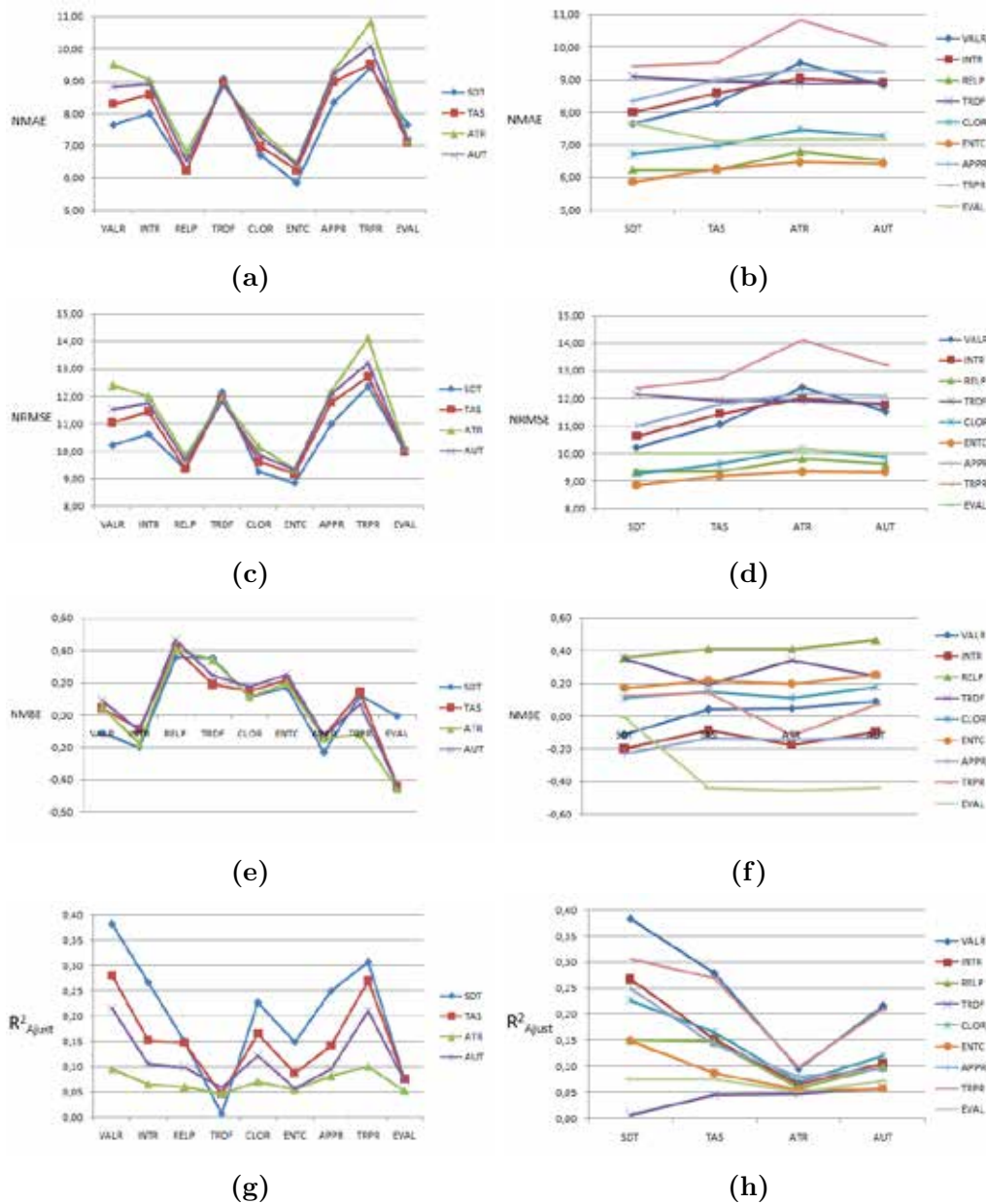


Figura 8.4: Ajuste global de modelos de regresión de la severidad global correspondientes a los modelos teóricos para cada sub-escala. SDT = Modelo SDT, TAS = Modelo TAS, ATR = Modelo Atribucional y AUT = Modelo de Auto-teorías.

Efectos de los modelos.

En esta sección se describen los efectos encontrados de las variables explicativas sobre la severidad global. El patrón de los efectos encontrados es relevante para evaluar la hipótesis central de esta investigación. También, nos

ayuda a evaluar en que medida las diferencias en severidad se pueden explicar mediante los modelos teóricos propuestos; y por tanto, en que medida aplican “hipótesis de validez” (R. Ames y Lau, 1979) de acuerdo a las perspectivas teóricas consideradas.

Adicionalmente, el patrón de efectos informa sobre el proceso de respuesta de los estudiantes, en relación a diferencias que pueden existir entre las dimensiones de la calidad en la enseñanza y respecto a las variables socio-cognitivas consideradas. Con el propósito de clarificar la interpretación de los resultados presentados, se describirá brevemente las escalas utilizadas para definir las variables y los efectos en el modelo.

Los coeficientes de regresión parcial correspondientes a cada variable explicativa (pendientes en el modelo de regresión múltiple) corresponden al efecto sobre la variable respuesta producido por un incremento en una unidad de la variable explicativa (1 logits). Para las variables explicativas, 1 logits representa entre un 5 % (para MOTIE) y 12 % (LOCUE) del rango de la variable. Por tanto, es una unidad que convenientemente expresa el efecto de las variables respuesta sobre las explicativas. Esto significa que la unidad es adecuada para describir diferencias que realmente podrían observarse.

Por otra parte, las unidades de las variables respuestas (severidad global), medida para cada una de las sub-escalas, representa en promedio un 5 % del rango de variación de cada variable, que también tiene como unidad 1 logits. Un incremento de 0,70 logits representa un tamaño de efecto superior al mínimo valor considerado de significación práctica, mientras que un cambio en 1 logits representa un tamaño de efecto considerado de magnitud aproximadamente moderada (Ferguson, 2009). La observación previa se relaciona con una orientación general propuesta por Ferguson (2009), no obstante, resulta un punto de partida útil para la interpretación de la significación práctica de un efecto, que puede refinarse con consideraciones más detalladas en cada caso.

El marco de referencia para medir la severidad global se definió tomando como centro u origen la dificultad promedio de los criterios de evaluación utilizados (ítems). Además, la severidad global fue centrada en ese valor, esto es se hizo coincidir con su media. Por tanto, una severidad igual a 0 logits es una severidad que tiene una medida igual a la dificultad promedio de los criterios de evaluación, y adicionalmente, coincide con la severidad global media. Por otra parte, al definir el marco de referencia para las medidas de las variables socio-cognitivas, la media de estas variables no se hizo coincidir con el origen.

Si bien, como antes, el origen del marco de referencia coincide también con la dificultad media de los ítems. Por tanto, una medida de 0 logits en las variables socio-cognitivas no coincide con la media de dichas variables, pero sí, con la media de la dificultad de los ítems.

Efectos en el modelo SDT.

En la Tabla 8.11 se muestran los efectos estimados mediante regresión robusta para las variables MOTIE y VALUE en el modelo de regresión SDT. Los efectos encontrados para MOTIE y VALUE sobre la severidad global para las sub-escalas fueron estadísticamente significativos, excepto, para el efecto de VALUE correspondiente a la escala TRPR. Por tanto, los efectos encontrados resultaron altamente consistentes con la hipótesis de investigación.

Adicionalmente, los efectos de MOTIE y VALUE fueron significativamente menores que cero para todas las escalas, excepto para la escala TRDR, donde el efecto de VALUE no fue estadísticamente significativo y el efecto de MOTIE fue significativamente mayor que cero. Por tanto, para 8/9 de las sub-escalas consideradas los efectos de las variables MOTIE y VALUE coinciden con las predicciones teóricas realizadas a partir del modelo SDT. Esto es, los efectos observados en el modelo de regresión presentan un alto nivel de acuerdo con las predicciones teóricas realizadas mediante el modelo SDT.

Observar que los efectos de MOTIE y VALUE sobre la severidad global, no solo son en general negativos, sino que suelen ser de un tamaño que puede considerarse, en términos generales, de significación práctica. Por ejemplo, para la sub-escala MOTIE, un incremento en 1 logits en la medida de MOTIE, se corresponde con una disminución en la severidad (-0,37 logits) al evaluar la dimensión “Valor y Utilidad” medida mediante la sub-escala VALR. A este efecto corresponde un valor de $R^2_{Parcial} = 0,12$, que indica un efecto de significación práctica, y además, próximo al valor 0,13, a partir del cual se interpreta que existe un tamaño de efecto moderado (J. Cohen, 1992).

Observar también, que en este caso un cambio en 1 logits, relativamente pequeño, en cada variable explicativa (MOTIE y VALUE) representa una reducción en la severidad global (-0,70 logits), que representa un valor mayor al de mínima significación práctica (Ferguson, 2009). Como se describió previamente, en la sección 8.3.4 el tamaño de efecto global encontrado para el modelo de regresión SDT fue en 7/9 sub-escalas moderado o elevado. Esto está vin-

culado con los signos y magnitudes de los efectos de MOTIE y VALUE. En el caso de la sub-escala TRPR (“Tareas Propuestas”) el efecto sobre la severidad global correspondiente al incremento en 1 logits en las variables explicativas es aún mayor (-0,83 logits).

Tabla 8.11: Modelo SDT.

| | | Valor | S.E. | t | p | pp _{erm} | BCa.CI _{2,5%} | BCa.CI _{97,5%} |
|-------|-------------|-------|------|--------|------|-------------------|------------------------|-------------------------|
| VALRE | (Intercept) | 0,76 | 0,05 | 16,28 | 0,00 | 0,00 | 0,67 | 0,86 |
| | MOTIE | -0,37 | 0,02 | -17,80 | 0,00 | 0,00 | -0,42 | -0,32 |
| | VALUE | -0,33 | 0,02 | -16,83 | 0,00 | 0,00 | -0,38 | -0,29 |
| | MOTIE:VALUE | 0,02 | 0,00 | 3,39 | 0,00 | 0,00 | 0,00 | 0,03 |
| INTRE | (Intercept) | 0,72 | 0,05 | 15,07 | 0,00 | 0,00 | 0,63 | 0,82 |
| | MOTIE | -0,34 | 0,02 | -16,34 | 0,00 | 0,00 | -0,40 | -0,30 |
| | VALUE | -0,17 | 0,02 | -8,59 | 0,00 | 0,00 | -0,22 | -0,12 |
| | MOTIE:VALUE | 0,01 | 0,00 | 1,69 | 0,09 | 0,09 | -0,01 | 0,02 |
| RELPE | (Intercept) | 0,82 | 0,06 | 12,76 | 0,00 | 0,00 | 0,69 | 0,97 |
| | MOTIE | -0,35 | 0,03 | -12,29 | 0,00 | 0,00 | -0,43 | -0,28 |
| | VALUE | -0,29 | 0,03 | -10,77 | 0,00 | 0,00 | -0,36 | -0,22 |
| | MOTIE:VALUE | 0,02 | 0,01 | 3,12 | 0,00 | 0,00 | 0,21 | 0,04 |
| TRDFE | (Intercept) | -0,03 | 0,05 | -0,64 | 0,53 | 0,43 | -0,15 | 0,06 |
| | MOTIE | 0,07 | 0,02 | 3,02 | 0,00 | 0,00 | 0,02 | 0,13 |
| | VALUE | 0,00 | 0,02 | -0,18 | 0,85 | 0,85 | -0,06 | 0,05 |
| | MOTIE:VALUE | 0,00 | 0,01 | -0,57 | 0,57 | 0,57 | -0,02 | 0,02 |
| CLORE | (Intercept) | 0,86 | 0,06 | 14,93 | 0,00 | 0,00 | 0,75 | 0,97 |
| | MOTIE | -0,38 | 0,03 | -15,04 | 0,00 | 0,00 | -0,44 | -0,32 |
| | VALUE | -0,26 | 0,02 | -11,03 | 0,00 | 0,00 | -0,33 | -0,21 |
| | MOTIE:VALUE | 0,02 | 0,01 | 3,63 | 0,00 | 0,00 | 0,00 | 0,04 |
| ENTCE | (Intercept) | 0,67 | 0,05 | 12,90 | 0,00 | 0,00 | 0,57 | 0,79 |
| | MOTIE | -0,34 | 0,02 | -14,76 | 0,00 | 0,00 | -0,39 | -0,28 |
| | VALUE | -0,18 | 0,02 | -8,33 | 0,00 | 0,00 | -0,23 | -0,13 |
| | MOTIE:VALUE | 0,01 | 0,00 | 2,83 | 0,01 | 0,01 | -0,14 | 0,03 |
| APPRE | (Intercept) | 0,83 | 0,05 | 15,48 | 0,00 | 0,00 | 0,74 | 0,96 |
| | MOTIE | -0,35 | 0,02 | -14,61 | 0,00 | 0,00 | -0,40 | -0,29 |
| | VALUE | -0,23 | 0,02 | -10,31 | 0,00 | 0,00 | -0,30 | -0,19 |
| | MOTIE:VALUE | 0,01 | 0,01 | 2,32 | 0,02 | 0,02 | 0,00 | 0,03 |
| TRPRE | (Intercept) | 1,03 | 0,07 | 15,45 | 0,00 | 0,00 | 0,91 | 1,17 |
| | MOTIE | -0,46 | 0,03 | -15,48 | 0,00 | 0,00 | -0,53 | -0,39 |
| | VALUE | -0,37 | 0,03 | -13,51 | 0,00 | 0,00 | -0,45 | -0,33 |
| | MOTIE:VALUE | 0,02 | 0,01 | 3,30 | 0,00 | 0,02 | 0,00 | 0,04 |
| EVALE | (Intercept) | 0,31 | 0,04 | 8,34 | 0,00 | 0,00 | 0,24 | 0,39 |
| | MOTIE | -0,23 | 0,02 | -13,98 | 0,00 | 0,00 | -0,27 | -0,19 |
| | VALUE | -0,07 | 0,02 | -4,26 | 0,00 | 0,00 | -0,10 | -0,03 |
| | MOTIE:VALUE | 0,00 | 0,00 | 1,22 | 0,22 | 0,00 | -0,01 | 0,02 |

Efectos en el modelo TAS.

En la Tabla 8.12 se muestran los efectos estimados mediante regresión robusta para las variables AUTE y ERESE en el modelo de regresión TAS. Los efectos encontrados para AUTE y ERESE sobre la severidad global para las sub-escalas fueron estadísticamente significativos para las nueve sub-escalas consideradas. Por tanto, los efectos encontrados resultaron altamente consistentes con la hipótesis central de esta investigación.

Adicionalmente, los efectos de AUTEE y ERESE fueron significativamente menores que cero para todas las escalas, excepto el efecto de la variable AUTEE sobre la severidad global TRDFE ($\text{valor} = 0,22$ logits; $BCa.CI_{95\%} = [0,15; 0,31]$). Por tanto, para todas las sub-escalas consideradas los efectos de las variables AUTEE y ERESE coinciden con las predicciones teóricas realizadas a partir del modelo TAS, salvo, el efecto de AUTEE sobre la variable TRDFE. Esto es, los efectos observados en el modelo de regresión presentan un alto nivel de acuerdo con las predicciones teóricas realizadas mediante el modelo TAS.

Observar que los efectos de AUTEE y ERESE sobre la severidad global, no solo son en general negativos, sino que suelen ser de un tamaño que puede considerarse de significación práctica, y mayores a los encontrados en el modelo de regresión SDT. Por ejemplo, para la variable ERESE, un incremento en 1 logits en su valor, se corresponde con una disminución en la severidad (-0,56 logits) al evaluar la dimensión “Tareas Propuestas” medida mediante la sub-escala TRPR. A este efecto corresponde un valor de $R_{\text{Parcial}}^2 = 0,08$, que supone un tamaño de efecto de significación práctica, de acuerdo al criterio general propuesto por [Ferguson \(2009\)](#) y los puntos de corte sugeridos por [J. Cohen \(1992\)](#).

Observar también, que en este caso un cambio en 1 logits, relativamente pequeño, en cada variable explicativa (AUTEE y ERESE) representa una reducción en la severidad global (TRPRE) equivalente a -1,02 logits, lo que representa un tamaño de efecto moderado, mayor al de mínima significación práctica ([Ferguson, 2009](#)). Como se describió previamente, en la sección [8.3.4](#) el tamaño de efecto global encontrado para el modelo de regresión TAS fue en cinco de las nueve sub-escalas moderado o elevado. Esto, al igual que para el modelo de regresión SDT, está relacionado con la coincidencia de signos y las magnitudes de los efectos de cada una de las variables explicativas.

Tabla 8.12: Modelo TAS.

| | | Valor | S.E. | t | p | pp _{erm} | BCa.CI _{2,5%} | BCa.CI _{97,5%} |
|-------|-------------|-------|------|--------|------|-------------------|------------------------|-------------------------|
| VALRE | (Intercept) | 0,68 | 0,05 | 13,06 | 0,00 | 0,00 | 0,57 | 0,78 |
| | AUTEE | -0,35 | 0,03 | -10,57 | 0,00 | 0,00 | -0,43 | -0,28 |
| | ERESE | -0,47 | 0,03 | -16,66 | 0,00 | 0,00 | -0,55 | -0,40 |
| | AUTEE:ERESE | 0,04 | 0,01 | 4,38 | 0,00 | 0,00 | 0,01 | 0,06 |
| INTRE | (Intercept) | 0,54 | 0,05 | 10,03 | 0,00 | 0,00 | 0,44 | 0,65 |
| | AUTEE | -0,19 | 0,03 | -5,48 | 0,00 | 0,00 | -0,27 | -0,12 |
| | ERESE | -0,35 | 0,03 | -11,92 | 0,00 | 0,00 | -0,42 | -0,28 |
| | AUTEE:ERESE | 0,03 | 0,01 | 3,48 | 0,00 | 0,00 | 0,01 | 0,06 |
| RELPE | (Intercept) | 0,78 | 0,07 | 11,51 | 0,00 | 0,00 | 0,65 | 0,92 |
| | AUTEE | -0,34 | 0,04 | -7,86 | 0,00 | 0,00 | -0,44 | -0,25 |
| | ERESE | -0,51 | 0,04 | -13,80 | 0,00 | 0,00 | -0,61 | -0,43 |
| | AUTEE:ERESE | 0,06 | 0,01 | 5,45 | 0,00 | 0,00 | 0,03 | 0,09 |
| TRDFE | (Intercept) | -0,23 | 0,05 | -4,22 | 0,00 | 0,00 | -0,34 | -0,12 |
| | AUTEE | 0,22 | 0,04 | 6,29 | 0,00 | 0,00 | 0,15 | 0,31 |
| | ERESE | -0,14 | 0,03 | -4,52 | 0,00 | 0,00 | -0,22 | -0,06 |
| | AUTEE:ERESE | 0,04 | 0,01 | 4,62 | 0,00 | 0,00 | 0,01 | 0,07 |
| CLORE | (Intercept) | 0,77 | 0,06 | 12,32 | 0,00 | 0,00 | 0,65 | 0,89 |
| | AUTEE | -0,34 | 0,04 | -8,54 | 0,00 | 0,00 | -0,44 | -0,26 |
| | ERESE | -0,43 | 0,03 | -12,71 | 0,00 | 0,00 | -0,52 | -0,36 |
| | AUTEE:ERESE | 0,05 | 0,01 | 5,41 | 0,00 | 0,00 | 0,03 | 0,08 |
| ENTCE | (Intercept) | 0,51 | 0,06 | 8,72 | 0,00 | 0,00 | 0,40 | 0,62 |
| | AUTEE | -0,20 | 0,04 | -5,42 | 0,00 | 0,00 | -0,29 | -0,12 |
| | ERESE | -0,33 | 0,03 | -10,39 | 0,00 | 0,00 | -0,41 | -0,26 |
| | AUTEE:ERESE | 0,04 | 0,01 | 3,80 | 0,00 | 0,00 | 0,01 | 0,06 |
| APPRE | (Intercept) | 0,64 | 0,06 | 10,46 | 0,00 | 0,00 | 0,53 | 0,77 |
| | AUTEE | -0,21 | 0,04 | -5,29 | 0,00 | 0,00 | -0,30 | -0,12 |
| | ERESE | -0,39 | 0,03 | -11,86 | 0,00 | 0,00 | -0,48 | -0,31 |
| | AUTEE:ERESE | 0,04 | 0,01 | 3,77 | 0,00 | 0,00 | 0,01 | 0,07 |
| TRPRE | (Intercept) | 1,07 | 0,07 | 15,15 | 0,00 | 0,00 | 0,93 | 1,21 |
| | AUTEE | -0,46 | 0,05 | -10,02 | 0,00 | 0,00 | -0,56 | -0,35 |
| | ERESE | -0,56 | 0,04 | -14,55 | 0,00 | 0,00 | -0,66 | -0,48 |
| | AUTEE:ERESE | 0,03 | 0,01 | 2,46 | 0,01 | 0,01 | 0,00 | 0,06 |
| EVALE | (Intercept) | 0,22 | 0,04 | 5,45 | 0,00 | 0,00 | 0,14 | 0,31 |
| | AUTEE | -0,15 | 0,03 | -5,71 | 0,00 | 0,00 | -0,21 | -0,10 |
| | ERESE | -0,13 | 0,02 | -6,06 | 0,00 | 0,00 | -0,18 | -0,08 |
| | AUTEE:ERESE | 0,01 | 0,01 | 2,00 | 0,05 | 0,04 | 0,00 | 0,03 |

Efectos en el modelo Atribucional.

En la Tabla 8.13 se muestran los efectos estimados mediante regresión robusta para las variables LOCUE, ESTAE y CONTE en el modelo de regresión Atribucional. Los efectos encontrados para LOCUE, ESTAE y CONTE sobre la severidad global para seis de las nueve sub-escalas fueron estadísticamente significativos. Para dos de las sub-escalas restantes (RELP y TRPR) los efectos de LOCUE y ESTAE fueron también estadísticamente significativos, pero, no fueron significativos para la variable CONTE. Por otra parte, ninguno de los efectos de las variables LOCUE, ESTAE y CONTE resultaron estadísticamente significativos sobre la variable TRDFE medida con la sub-escala TRDF.

Para una correcta interpretación de los análisis previos son importantes algunas observaciones. La primera, refiere a la dimensión medida por la sub-

escala TRDF (“Carga de Trabajo y Dificultad”). La sub-escala TRDF contiene ítems que refieren a la dificultad, carga de trabajo y el ritmo de la enseñanza. Una dificultad inherente de esta dimensión de la calidad en la enseñanza, se relaciona con el hecho que su efecto facilitador sobre el proceso de aprendizaje de los estudiantes tiene una forma de “U invertida” o de “Campana” (Atkinson, 1957; Kukla, 1972; Schunk y cols., 2014; Weiner, 1992).

Así, para cada estudiante hay un nivel óptimo que implica un máximo efecto facilitador. Por debajo, o por encima de dicho nivel óptimo, el efecto facilitador de la enseñanza es menor. Este patrón de “U invertida”, podría explicar la no significación estadística de los efectos encontrados de las variables LOCUE, ESTAE y CONTE sobre TRDFE. Por ejemplo, consideremos dos valores de TRDFE, un valor bajo y uno alto. Ambos valores podrían relacionarse con un bajo efecto facilitador de la enseñanza, mientras que ciertos valores intermedios de TRDFE podrían relacionarse con un mayor efecto facilitador.

Una segunda observación se refiere a la dimensión de controlabilidad y la medida obtenida (CONTE). Si se asume que todas las causas externas son incontrolables, la dimensión de controlabilidad significa “controlable por mi”. Esto suele asumirse en la teoría de las atribuciones causales de Weiner. No obstante, Weiner (1985) reconoce la posibilidad que algunos sujetos perciban ciertas causas externas como “controlables por otros”.

Por ejemplo, la dificultad suele considerarse una causa externa, estable e incontrolable. No obstante, algunos sujetos podrían percibirla como externa, estable y controlable por otros. Por otra parte, los efectos sobre los sujetos de las atribuciones “no controlables” y “controlables por otros” podrían ser diferentes. Adicionalmente, si los efectos de atribuciones “controlables por mi” y “controlables por otros” fueran similares, tal vez, se podría hablar de “controlable por cualquiera” (Weiner, 1985).

Estas consideraciones fueron abordadas empíricamente por McAuley, Duncan, y Russell (1992). Estos autores aplicaron la técnica de análisis factorial confirmatorio, de datos provenientes de cuatro estudios, de la escala CDSII (“Causal Dimensions Scale”). A partir de estos análisis aportaron evidencias que sugieren que la dimensión de controlabilidad debería separarse en dos dimensiones (“control personal” y “control externo”). Adicionalmente, McAuley y cols. (1992) mostraron que el mejor ajuste entre los datos y el modelo, comparado con otras alternativas consideradas, se obtiene cuando se incluyen las siguientes dimensiones: locus de causalidad, estabilidad, control personal y

control externo.

En el presente trabajo la medida CONTE se obtuvo por un método indirecto. Las dimensiones de causalidad consideradas fueron las definidas por la teoría Atribucional de Weiner. Por tanto, causas que podrían considerarse bajo control externo, se trataron como incontrolables. No obstante, en el contexto de este estudio, esto parece no haber afectado la definición del efecto facilitador de la enseñanza en relación a la dimensión controlabilidad (CONTD). Un indicio de esto, es que las correlaciones estadísticamente significativas entre el efecto facilitador y las medidas de calidad en la enseñanza fueron todas positivas (resultados no presentados aquí). Esto es lo que se esperaría cuando se determina la correlación de dos medidas alternativas, validas, de la calidad en la enseñanza.

Por otra parte, los efectos de CONTE en la severidad global de todas las sub-escalas fue positivo, y estadísticamente significativo en seis de nueve de las sub-escalas. Adicionalmente, en el análisis correlacional, se encontró una correlación positiva estadísticamente significativa entre EVALE y CONTE. Por otra parte, la medida del componente de severidad CONTE[C9], obtenida mediante el modelo LLT-MFRM fue en general significativamente mayor que cero, por tanto, a medida que se incrementa el peso de este componente en la medida de severidad global, esta se incrementa (los estudiantes se tornan más severos).

No obstante, es importante considerar que el modelo Atribucional del proceso de respuesta a los SET plantea que a medida que un patrón atribucional se aleja de un patrón atribucional II y se aproxima a un patrón atribucional I, la severidad de los estudiantes al evaluar la enseñanza debería disminuir. Este modelo considera patrones de dimensiones de causalidad, no sus efectos aislados. Esto es así, pues, las dimensiones de causalidad están estrechamente relacionadas entre sí, y permiten caracterizar atribuciones a causas específicas y sus efectos emocionales, cognitivos y comportamentales, cuando se consideran en conjunto.

La condición impuesta, esto es, efectos negativos de las variables LOCUE, ESTAE y CONTE, en relación a los efectos en el modelo de regresión atribucional, es consistente en términos generales, con el modelo atribucional. No obstante, la condición de que todos los efectos sean negativos en el modelo de regresión atribucional es condición suficiente, pero, no necesaria para especificar en el modelo de regresión atribucional, lo que expresa el modelo atribucional

del proceso de respuesta. Para verificar con mayor precisión si las predicciones teóricas del modelo atribucional son consistentes con los resultados obtenidos, a continuación se determina en varios casos los cambios encontrados en la severidad, cuando el patrón atribucional se aproxima al patrón atribucional I y se aleja del patrón atribucional II.

Por tanto, para clarificar el efecto global que tendrían las tres variables LOCUE, ESTAE y CONTE, es necesario considerar el patrón de efectos en su conjunto. Por ejemplo, consideremos un cambio en las atribuciones causales realizadas por un sujeto de un patrón cualitativamente similar a la dificultad (locus externo, estable, incontrolable) a un patrón similar al esfuerzo (locus interno, inestable, controlable). Esto es, el cambio realizado implica alejarse del patrón atribucional II y aproximarse al patrón atribucional I.

Para simplificar el análisis, consideremos que la diferencia entre los dos patrones es de 1 logits. Esto es, por ejemplo, al cambiar de patrón pasamos a un locus más interno en una magnitud de 1 logits. Si el primer patrón de valores de las variables LOCUE, ESTAE y CONTE corresponde en logits a (0,0,0), el segundo patrón correspondería a (1,1,1). Si la magnitud del cambio fuera distinto de 1 logits, los resultados serían cualitativamente iguales. Por ejemplo, si fueron 5 logits en vez de 1 logits, el cambio en la severidad sería en el mismo sentido, pero, de mayor magnitud.

Si ahora se determina el cambio esperado en severidad (CES) de acuerdo al modelo de regresión atribucional para la variable VALRE, se encuentra que la reducción en la severidad global es igual a $-0,96$ logits. Para las otras sub-escalas los valores obtenidos en logits son: $CES_{INTRE} = -0,52$, $CES_{RELPE} = -1,04$, $CES_{TRDFE} = 0,16$, $CES_{CLORE} = -0,93$, $CES_{ENTC} = -0,48$, $CES_{APPR} = -0,59$, $CES_{TRPR} = -1,57$ y $CES_{TRPR} = -0,19$. Este análisis muestra que los efectos encontrados en el modelo de regresión atribucional son altamente consistentes con el modelo atribucional del proceso de respuesta a los SET, cuando los cambios en cada una de las dimensiones son similares en magnitud.

No obstante, si fueran posibles, cambios mucho mayores en la controlabilidad que en las otras dimensiones, en algunos casos se encontrarían resultados no consistentes con el modelo atribucional. Sin embargo, esto no es posible para el efecto de cambios en los patrones atribucionales sobre la variable TRPRE. De acuerdo al modelo de regresión atribucional para dicha variable, los efectos de los patrones (LOCUE, ESTAE, CONTE) siempre serían consistentes con las

predicciones teóricas. En otros casos, el cambio en CONTE debería ser entre 2 y 4 veces mayor para que se produzca una discrepancia con el modelo atribucional. Debido a que las dimensiones de causalidad están estrechamente relacionadas y no cualquier combinación de valores es posible, estos casos de desviaciones respecto al modelo atribucional parecen poco plausibles.

Por tanto, los efectos encontrados resultaron altamente consistentes con la hipótesis central de esta investigación. Por otra parte, los efectos de las variables LOCUE y ESTAE son altamente consistentes con los predichos por el modelo Atribucional. Claramente se constata que el efecto de CONTE es positivo, y por tanto, contribuye a una mayor severidad. No obstante, las diferencias en la severidad entre patrones atribucionales parecen ser ampliamente consistentes con las predichas por el modelo Atribucional.



Tabla 8.13: Modelo Atribucional.

| | | Valor | S.E. | t | p | pp _{perm} | BCa.CI _{2,5%} | BCa.CI _{97,5%} |
|-------|-------------|-------|------|-------|------|--------------------|------------------------|-------------------------|
| VALRE | (Intercept) | 0,77 | 0,07 | 10,44 | 0,00 | 0,00 | 0,62 | 0,93 |
| | LOCUE | -0,99 | 0,20 | -4,96 | 0,00 | 0,00 | -1,41 | -0,61 |
| | ESTAE | -0,38 | 0,07 | -5,02 | 0,00 | 0,00 | -0,61 | -0,24 |
| | CONTE | 0,35 | 0,18 | 1,94 | 0,05 | 0,05 | -0,07 | 0,66 |
| | LOCUE:ESTAE | -0,20 | 0,14 | -1,43 | 0,15 | 0,15 | -0,56 | 0,10 |
| | LOCUE:CONTE | 0,05 | 0,03 | 1,80 | 0,07 | 0,07 | 0,00 | 0,14 |
| | ESTAE:CONTE | 0,21 | 0,08 | 2,55 | 0,01 | 0,01 | 0,03 | 0,44 |
| INTRE | (Intercept) | 0,68 | 0,07 | 9,98 | 0,00 | 0,00 | 0,55 | 0,82 |
| | LOCUE | -0,99 | 0,19 | -5,31 | 0,00 | 0,00 | -1,40 | -0,66 |
| | ESTAE | -0,26 | 0,07 | -3,79 | 0,00 | 0,00 | -0,45 | -0,13 |
| | CONTE | 0,56 | 0,17 | 3,39 | 0,00 | 0,00 | 0,24 | 0,87 |
| | LOCUE:ESTAE | 0,03 | 0,13 | 0,20 | 0,84 | 0,84 | -0,24 | 0,36 |
| | LOCUE:CONTE | 0,05 | 0,03 | 1,74 | 0,08 | 0,08 | 0,01 | 0,12 |
| | ESTAE:CONTE | 0,09 | 0,08 | 1,22 | 0,22 | 0,22 | -0,09 | 0,26 |
| RELPE | (Intercept) | 1,02 | 0,09 | 11,41 | 0,00 | 0,00 | 0,82 | 1,19 |
| | LOCUE | -1,05 | 0,24 | -4,34 | 0,00 | 0,00 | -1,61 | -0,58 |
| | ESTAE | -0,40 | 0,09 | -4,42 | 0,00 | 0,00 | -0,69 | -0,18 |
| | CONTE | 0,34 | 0,21 | 1,60 | 0,11 | 0,11 | -0,10 | 0,80 |
| | LOCUE:ESTAE | -0,18 | 0,17 | -1,12 | 0,26 | 0,25 | -0,61 | 0,24 |
| | LOCUE:CONTE | 0,04 | 0,03 | 1,25 | 0,21 | 0,20 | -0,02 | 0,13 |
| | ESTAE:CONTE | 0,21 | 0,10 | 2,08 | 0,04 | 0,03 | -0,10 | 0,47 |
| TRDFE | (Intercept) | -0,02 | 0,07 | -0,25 | 0,80 | 0,77 | -0,15 | 0,10 |
| | LOCUE | 0,07 | 0,18 | 0,38 | 0,70 | 0,69 | -0,35 | 0,45 |
| | ESTAE | -0,10 | 0,07 | -1,47 | 0,14 | 0,14 | -0,26 | 0,08 |
| | CONTE | 0,18 | 0,16 | 1,10 | 0,27 | 0,26 | -0,15 | 0,58 |
| | LOCUE:ESTAE | -0,08 | 0,12 | -0,65 | 0,51 | 0,51 | -0,39 | 0,25 |
| | LOCUE:CONTE | -0,01 | 0,03 | -0,25 | 0,80 | 0,81 | -0,06 | 0,06 |
| | ESTAE:CONTE | 0,10 | 0,07 | 1,35 | 0,18 | 0,18 | -0,10 | 0,30 |
| CLORE | (Intercept) | 1,01 | 0,08 | 12,48 | 0,00 | 0,00 | 0,85 | 1,17 |
| | LOCUE | -1,20 | 0,22 | -5,46 | 0,00 | 0,00 | -1,71 | -0,79 |
| | ESTAE | -0,32 | 0,08 | -3,93 | 0,00 | 0,00 | -0,56 | -0,11 |
| | CONTE | 0,50 | 0,20 | 2,55 | 0,01 | 0,01 | 0,11 | 0,91 |
| | LOCUE:ESTAE | -0,25 | 0,15 | -1,67 | 0,09 | 0,09 | -0,66 | 0,10 |
| | LOCUE:CONTE | 0,07 | 0,03 | 2,15 | 0,03 | 0,03 | 0,01 | 0,15 |
| | ESTAE:CONTE | 0,27 | 0,09 | 2,97 | 0,00 | 0,00 | 0,06 | 0,54 |
| ENTCE | (Intercept) | 0,71 | 0,07 | 9,57 | 0,00 | 0,00 | 0,56 | 0,86 |
| | LOCUE | -0,90 | 0,20 | -4,46 | 0,00 | 0,00 | -1,36 | -0,57 |
| | ESTAE | -0,23 | 0,08 | -3,05 | 0,00 | 0,00 | -0,42 | -0,06 |
| | CONTE | 0,67 | 0,18 | 3,77 | 0,00 | 0,00 | 0,34 | 1,03 |
| | LOCUE:ESTAE | -0,24 | 0,14 | -1,75 | 0,08 | 0,08 | -0,60 | 0,11 |
| | LOCUE:CONTE | -0,02 | 0,03 | -0,79 | 0,43 | 0,42 | -0,07 | 0,05 |
| | ESTAE:CONTE | 0,24 | 0,08 | 2,85 | 0,00 | 0,00 | 0,02 | 0,46 |
| APPRE | (Intercept) | 0,82 | 0,08 | 10,66 | 0,00 | 0,00 | 0,66 | 0,98 |
| | LOCUE | -0,95 | 0,21 | -4,53 | 0,00 | 0,00 | -1,41 | -0,56 |
| | ESTAE | -0,33 | 0,08 | -4,22 | 0,00 | 0,00 | -0,54 | -0,13 |
| | CONTE | 0,73 | 0,19 | 3,91 | 0,00 | 0,00 | 0,37 | 1,12 |
| | LOCUE:ESTAE | -0,18 | 0,14 | -1,29 | 0,19 | 0,20 | -0,58 | 0,17 |
| | LOCUE:CONTE | -0,03 | 0,03 | -0,95 | 0,34 | 0,35 | -0,08 | 0,04 |
| | ESTAE:CONTE | 0,17 | 0,09 | 1,94 | 0,05 | 0,05 | -0,05 | 0,41 |
| TRPRE | (Intercept) | 1,30 | 0,10 | 13,39 | 0,00 | 0,00 | 1,14 | 1,51 |
| | LOCUE | -0,80 | 0,26 | -3,06 | 0,00 | 0,00 | -1,45 | -0,32 |
| | ESTAE | -0,69 | 0,10 | -7,01 | 0,00 | 0,00 | -1,03 | -0,49 |
| | CONTE | -0,01 | 0,23 | -0,05 | 0,10 | 0,97 | -0,57 | 0,44 |
| | LOCUE:ESTAE | -0,22 | 0,18 | -1,20 | 0,23 | 0,23 | -0,63 | 0,29 |
| | LOCUE:CONTE | 0,01 | 0,04 | 0,30 | 0,77 | 0,77 | -0,06 | 0,12 |
| | ESTAE:CONTE | 0,14 | 0,11 | 1,25 | 0,21 | 0,22 | -0,18 | 0,39 |
| EVALE | (Intercept) | 0,29 | 0,05 | 5,73 | 0,00 | 0,00 | 0,20 | 0,39 |
| | LOCUE | -0,37 | 0,14 | -2,76 | 0,01 | 0,01 | -0,65 | -0,10 |
| | ESTAE | -0,17 | 0,05 | -3,47 | 0,00 | 0,00 | -0,30 | -0,06 |
| | CONTE | 0,42 | 0,12 | 3,50 | 0,00 | 0,00 | 0,18 | 0,67 |
| | LOCUE:ESTAE | -0,13 | 0,09 | -1,42 | 0,16 | 0,16 | -0,37 | 0,10 |
| | LOCUE:CONTE | -0,05 | 0,02 | -2,81 | 0,00 | 0,01 | -0,10 | -0,01 |
| | ESTAE:CONTE | 0,11 | 0,06 | 1,91 | 0,06 | 0,06 | -0,03 | 0,24 |

Efectos en el modelo de Auto-teorías.

En la Tabla 8.14 se muestran los efectos estimados mediante regresión robusta para las variables AUTEe y ELEME en el modelo de regresión de Auto-teorías. Los efectos encontrados para AUTEe y ELEME sobre la severidad global para las sub-escalas fueron estadísticamente significativos, excepto, para el efecto de ELEME correspondiente a la escala TRPR ($Valor = -0,03$ logits, $p_{Perm} = 0,36$, $BCa.CI_{95\%} = [-0,11; 0,03]$). Por tanto, los efectos encontrados resultaron altamente consistentes con la hipótesis central de esta investigación.

Adicionalmente, los efectos de AUTEe y ELEME fueron significativamente menores que cero para siete de las nueve sub-escalas. Para las dos sub-escalas restantes una de las variables explicativas tiene un efecto negativo y la otra positivo. En ambos casos los efectos positivos son bastante pequeños. Para la sub-escala TRDF el efecto de la variable AUTEe resultó significativamente mayor que cero ($Valor = 0,08$ logits, $p_{Perm} = 0,00$, $BCa.CI_{95\%} = [0,00; 0,13]$). A su vez, para la sub-escala EVAL el efecto de la variable ELEME resultó significativamente mayor que cero ($Valor = 0,06$ logits, $p_{Perm} = 0,01$, $BCa.CI_{95\%} = [0,01; 0,10]$). En este caso, los efectos encontrados muestran también una buena consistencia entre los efectos observados y los predichos a partir del modelo de Auto-teorías.

Sumario: Patrón de efectos encontrados y validez de constructo de las sub-escalas.

En la Figura 8.5 se presenta una síntesis gráfica de los resultados obtenidos en esta sección. Aquí la hipótesis central de esta investigación se evaluó desde cuatro perspectivas teóricas distintas. Adicionalmente, se analizaron los efectos encontrados con cuatro modelos teóricos del proceso de respuesta que fueron propuestos en el capítulo 4.

Estos modelos representan especificaciones o restricciones de la hipótesis central de esta investigación, que definen “hipótesis de validez” de la evaluación estudiantil de la enseñanza. Un buen acuerdo con los modelos teóricos considerados, significan que la información obtenida a partir de las evaluaciones estudiantiles de la enseñanza y proceso de aprendizaje aportan información válida. Así, a pesar de que existe un elevado efecto del evaluador, lo que implica un alto grado de subjetividad, no obstante, la información recabada de

esta forma es válida y útil para comprender la enseñanza, el aprendizaje y sus inter-relaciones.

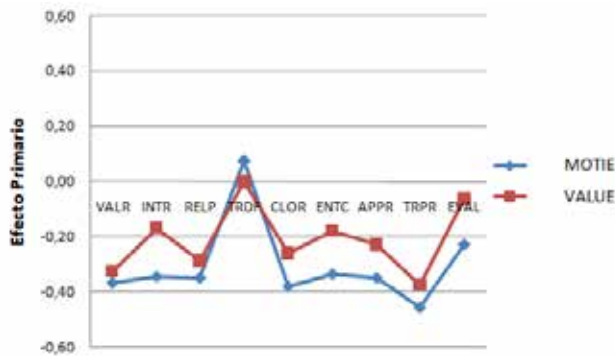
Los resultados encontrados desde las cuatro perspectivas teóricas consideradas son ampliamente consistentes con la hipótesis central de esta investigación. Adicionalmente, se encontró un alto nivel de consistencia de las predicciones teóricas de los modelos considerados con los efectos encontrados. Esto, significa que en conjunto se encontraron evidencias importantes de validez de constructo del cuestionario de evaluación de calidad en la enseñanza y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje. No obstante, los resultados analizados sugieren algunas limitaciones.

En la Figura 8.5 se aprecia una desviación sistemática entre el patrón esperado de efectos para las medidas de severidad global en la sub-escala TRDF. Los resultados presentados en esta sección aportaron algunos indicios para comprender diferencias que existen en el proceso de respuesta a esta sub-escala.

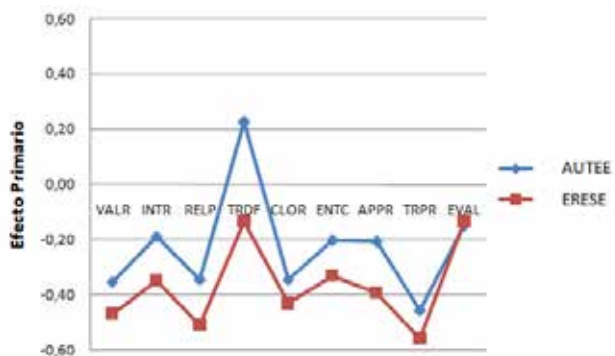
Adicionalmente, los hallazgos presentados en esta sección sugieren una revisión de la medida CONTE y de la sub-escala empleada para su medición. Si bien el patrón encontrado en los efectos de CONTE sobre la severidad global parecen no ser inconsistentes con el modelo Atribucional, no es claro que la diferencia encontrada, con respecto a las otras variables (LOCUE y ESTAE) reflejen una diferencia relevante para el constructo analizado (véase Figura 8.5).

Tabla 8.14: Modelo de Auto-teorías.

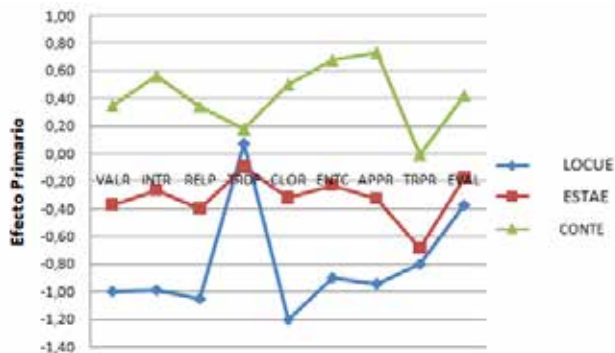
| | | Valor | S.E. | t | p | pPerm | BCa.CI _{2,5%} | BCa.CI _{97,5%} |
|-------|-------------|-------|------|--------|------|-------|------------------------|-------------------------|
| VALRE | (Intercept) | 0,78 | 0,06 | 12,31 | 0,00 | 0,00 | 0,64 | 0,93 |
| | AUTEE | -0,63 | 0,03 | -20,94 | 0,00 | 0,00 | -0,72 | -0,57 |
| | ELEME | -0,06 | 0,03 | -2,07 | 0,04 | 0,04 | -0,15 | 0,00 |
| | AUTEE:ELEME | 0,01 | 0,01 | 0,52 | 0,61 | 0,61 | -0,02 | 0,05 |
| INTRE | (Intercept) | 0,67 | 0,06 | 10,78 | 0,00 | 0,00 | 0,56 | 0,81 |
| | AUTEE | -0,41 | 0,03 | -14,02 | 0,00 | 0,00 | -0,49 | -0,34 |
| | ELEME | -0,09 | 0,03 | -3,05 | 0,00 | 0,00 | -0,17 | -0,03 |
| | AUTEE:ELEME | 0,03 | 0,01 | 2,10 | 0,04 | 0,03 | -0,01 | 0,06 |
| RELPE | (Intercept) | 0,96 | 0,08 | 12,19 | 0,00 | 0,00 | 0,78 | 1,11 |
| | AUTEE | -0,60 | 0,04 | -16,05 | 0,00 | 0,00 | -0,69 | -0,52 |
| | ELEME | -0,13 | 0,04 | -3,48 | 0,00 | 0,00 | -0,22 | -0,05 |
| | AUTEE:ELEME | 0,02 | 0,02 | 1,14 | 0,25 | 0,25 | -0,02 | 0,06 |
| TRDFE | (Intercept) | -0,17 | 0,06 | -2,75 | 0,01 | 0,00 | -0,28 | -0,03 |
| | AUTEE | 0,08 | 0,03 | 2,83 | 0,00 | 0,00 | 0,00 | 0,13 |
| | ELEME | -0,03 | 0,03 | -0,90 | 0,37 | 0,36 | -0,11 | 0,03 |
| | AUTEE:ELEME | 0,08 | 0,01 | 6,95 | 0,00 | 0,00 | 0,06 | 0,12 |
| CLORE | (Intercept) | 0,94 | 0,07 | 12,92 | 0,00 | 0,00 | 0,80 | 1,09 |
| | AUTEE | -0,59 | 0,03 | -17,02 | 0,00 | 0,00 | -0,67 | -0,52 |
| | ELEME | -0,12 | 0,03 | -3,40 | 0,00 | 0,00 | -0,21 | -0,04 |
| | AUTEE:ELEME | 0,04 | 0,01 | 2,62 | 0,01 | 0,01 | 0,00 | 0,08 |
| ENTCE | (Intercept) | 0,63 | 0,07 | 9,38 | 0,00 | 0,00 | 0,52 | 0,79 |
| | AUTEE | -0,41 | 0,03 | -12,88 | 0,00 | 0,00 | -0,50 | -0,35 |
| | ELEME | -0,09 | 0,03 | -2,75 | 0,01 | 0,01 | -0,19 | -0,04 |
| | AUTEE:ELEME | 0,03 | 0,01 | 2,46 | 0,01 | 0,01 | 0,01 | 0,08 |
| APPRE | (Intercept) | 0,78 | 0,07 | 10,96 | 0,00 | 0,00 | 0,65 | 0,93 |
| | AUTEE | -0,46 | 0,03 | -13,66 | 0,00 | 0,00 | -0,54 | -0,38 |
| | ELEME | -0,11 | 0,03 | -3,14 | 0,00 | 0,00 | -0,19 | -0,03 |
| | AUTEE:ELEME | 0,04 | 0,01 | 2,85 | 0,00 | 0,00 | 0,00 | 0,08 |
| TRPRE | (Intercept) | 1,23 | 0,08 | 14,95 | 0,00 | 0,00 | 1,09 | 1,39 |
| | AUTEE | -0,81 | 0,04 | -20,68 | 0,00 | 0,00 | -0,90 | -0,74 |
| | ELEME | -0,09 | 0,04 | -2,33 | 0,02 | 0,02 | -0,19 | -0,02 |
| | AUTEE:ELEME | 0,00 | 0,02 | -0,26 | 0,79 | 0,79 | -0,04 | 0,04 |
| EVALE | (Intercept) | 0,18 | 0,05 | 3,91 | 0,00 | 0,00 | 0,09 | 0,26 |
| | AUTEE | -0,25 | 0,02 | -11,42 | 0,00 | 0,00 | -0,31 | -0,21 |
| | ELEME | 0,06 | 0,02 | 2,66 | 0,01 | 0,01 | 0,01 | 0,10 |
| | AUTEE:ELEME | 0,01 | 0,01 | 1,23 | 0,22 | 0,21 | -0,01 | 0,04 |



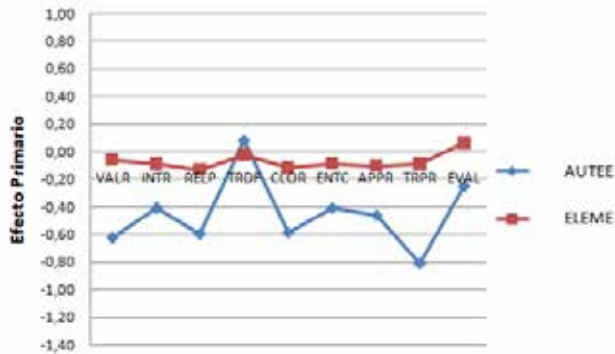
(a) Modelo SDT.



(b) Modelo TAS.



(c) Modelo Atribucional.



(d) Modelo de Auto-teorías.

Figura 8.5: Modelos teóricos del proceso de respuesta a los SETs y efectos primarios encontrados.

8.4. Discusión.

La investigación reportada en este capítulo se llevó a cabo con dos objetivos primarios. En primer lugar, evaluar en forma más directa la hipótesis central de esta investigación. En segundo lugar, aportar evidencias adicionales de validez de constructo del cuestionario de evaluación de la calidad en la enseñanza y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje.

Esto se realizó, concentrando el análisis en la relación entre la severidad global de los estudiantes al evaluar la calidad de la enseñanza y las medidas de los estudiantes en nueve variables socio-cognitivas: auto-eficacia para aprender, expectativas positivas de resultado, motivación intrínseca, percepción de elección, percepción de valor y utilidad, elección de metas, locus de causalidad, estabilidad y controlabilidad.

La aproximación adoptada presentó varias características singulares. Por un lado, el abordaje teórico implicó la aproximación desde varias perspectivas teóricas a los objetivos planteados. Así, la hipótesis central de esta investigación y la validez de constructo relacionada con los cuestionarios, explícitamente se analizó desde la perspectiva de cuatro modelos teóricos: el modelo SDT, el modelo TAS, el modelo Atribucional y el modelo de Auto-teorías. Estos modelos fueron propuestos y descriptos en el capítulo 4 de esta tesis. A partir de la hipótesis central de esta investigación y de los cuatro modelos teóricos se derivaron una serie de predicciones teóricas, que fueron detalladamente contrastadas con los patrones observados en los datos.

Por otro lado, la aproximación metodológica se fundamentó en la aplicación de diferentes abordajes metodológicos al mismo objeto de estudio. En primer lugar, se aplicó un modelo de medidas invariantes explicativo (LLT-MFRM) que permitió descomponer la medida de severidad global en una serie de componentes o factores de severidad (Engelhard Jr, 2013; M. Wilson y De Boeck, 2004). En segundo lugar, se realizó un análisis correlacional entre las medidas de calidad en la enseñanza y las medidas de las variables socio-cognitivas, para analizar las evidencias de validez convergente y discriminante (Albright y Malloy, 2000; Campbell y Fiske, 1959). En tercer lugar, se derivaron modelos de regresión para la variable severidad global, en los que la selección de las variables explicativas se fundamentó en los modelos teóricos propuestos en el capítulo 4.

Un aspecto importante de este trabajo es que se evaluó en forma sistemática

y detallada el grado de ajuste entre la información aportada y la base teórica racional subyacente. Este tipo de evidencia se ha considerado como la de mayor fuerza y valor informativo, y constituye un aporte clave a la comprensión de la validez de constructo (Cronbach, 1988; Kane, 1992; Messick, 1995).

Los resultados obtenidos son ampliamente consistentes con las predicciones teóricas realizadas y constituyen una evidencia fuerte para la hipótesis central de esta investigación.

En primer lugar, se encontró un ajuste aceptable entre los datos y el modelo LLT-MFRM, para cada una de las sub-escalas consideradas. Adicionalmente, a pesar de la importante reducción en el número de elementos de medición considerados en este modelo, respecto al modelo MFRM original, se encontró un ajuste similar. Esto sugiere que las restricciones lineales en la severidad, en el modelo LLT-MFRM, capturan razonablemente el papel de las variables socio-cognitivas en el proceso de respuesta de los estudiantes.

Adicionalmente, el patrón general observado en relación a los componentes de severidad, muestra que los componentes presentan un buen ajuste, en general con medidas significativamente distintas de cero, para cada una de las sub-escalas consideradas. A su vez, los componentes mostraron medidas consistentes con la medidas de cada sub-escala, como lo sugieren los valores de correlación punto medida (r_{pm}). Además, las medidas de los componentes de severidad mostraron una elevada confiabilidad como lo indicaron los valores obtenidos de los índices de confiabilidad de Rasch (SR , G y H).

Los resultados del análisis del patrón de correlaciones entre las medidas de severidad y las medidas de las variables socio-cognitivas, también fueron consistentes con la hipótesis central de esta investigación. Estos resultados muestran claramente que las medidas de severidad presentaron una correlación significativamente distinta de cero con las medidas de las variables socio-cognitivas de los estudiantes.

En forma similar, los efectos de las variables socio-cognitivas en los modelos de regresión, relacionados con los modelos teóricos del proceso de respuesta, fueron en general estadísticamente significativos.

En síntesis, los resultados obtenidos triangulando métodos y teorías, señalan consistentemente que las variables socio-cognitivas consideradas explican diferencias en la severidad de los estudiantes al evaluar la enseñanza.

En relación al segundo objetivo de esta investigación, aportar evidencias adicionales de validez de constructo, un aspecto clave abordado fue entender

en que medida el proceso de respuesta de los estudiantes es capaz de aportar información fidedigna y útil. Este problema se abordó en base a cuatro modelos teóricos que fueron propuestos y descritos en el capítulo 4. Estos modelos representan especificaciones o restricciones de la hipótesis central de esta investigación, que definen “hipótesis de validez” (R. Ames y Lau, 1979) de la evaluación estudiantil de la enseñanza. Un buen acuerdo con los modelos teóricos considerados, significan que la información obtenida a partir de las evaluaciones estudiantiles de la enseñanza y proceso de aprendizaje aportan información útil y fidedigna.

Los resultados encontrados indicaron un alto nivel de consistencia de las predicciones teóricas de los modelos considerados con los efectos encontrados. Esto, significa que en conjunto se encontraron evidencias importantes de validez de constructo del cuestionario de evaluación de calidad en la enseñanza y del cuestionario de evaluación socio-cognitiva del proceso de aprendizaje.

Por otra parte, los resultados encontrados muestran claramente que las variables socio-cognitivas consideradas serían importantes variables explicativas de la severidad global de los estudiantes al evaluar la calidad de la enseñanza. Además, en general tendrían un efecto negativo sobre la severidad global de los estudiantes. Estos resultados son consistentes con hallazgos encontrados en estudios previos (R. Ames y Lau, 1979; Grimes y cols., 2004; Svanum y Aigner, 2011; Wigfield y cols., 1997).

Por ejemplo, resultó claro que estudiantes con una motivación autónoma elevada, que presentan una motivación intrínseca alta y/o una elevada percepción del valor y utilidad de la enseñanza, serían menos severos que estudiantes con un nivel bajo en estas variables. En forma similar, estudiantes con un alto nivel de autoeficacia para aprender y expectativas de resultados positivos altas, serían también menos severos que otros estudiantes con bajos niveles en estas variables.

Adicionalmente, el enfoque teórico y metodológico en este capítulo, permitió detectar algunas posibles limitaciones en dos de las sub-escalas consideradas. Esto no había sido advertido, con claridad, con otras metodologías empleadas.

En primer lugar, los resultados encontrados en relación a la medida de controlabilidad de los estudiantes (CONTE) y la sub-escala empleada para su medición, sugieren la necesidad de una revisión e investigación más profunda de estas. Si bien el patrón encontrado en los efectos de CONTE sobre la severidad

global parecen no ser inconsistentes con el modelo Atritual, no es claro que la diferencia encontrada, con respecto a las otras variables (LOCUE y ESTAE) reflejen una diferencia relevante para el constructo analizado (veáse Figura 8.5).

Además, se encontró una desviación sistemática entre el patrón esperado de efectos para las medidas de severidad global en la sub-escala TRDF. Estas desviaciones podrían estar relacionadas con la relación existente entre la dimensión medida por esta sub-escala y su efecto facilitador del aprendizaje.

La sub-escala TRDF contiene ítems que refieren a la dificultad, carga de trabajo y el ritmo de la enseñanza. Una dificultad inherente de esta dimensión de la calidad en la enseñanza, se relaciona con el hecho que su efecto facilitador sobre el proceso de aprendizaje de los estudiantes tiene una forma de “U invertida” o de “Campana” (Atkinson, 1957; Kukla, 1972; Schunk y cols., 2014; Weiner, 1992). Se considera que esto puede ser una explicación plausible para la desviación sistemática observa respecto a la sub-escala TRDF.



Capítulo 9

Consideraciones finales.

9.1. Hipótesis de investigación y evidencias encontradas.

La hipótesis central que se propuso evaluar aquí expresa que: existen diferencias individuales en cómo los estudiantes evalúan la enseñanza, relacionadas con variables que median la interpretación cognitiva de las señales utilizadas para evaluar la enseñanza; o bien, que explican diferencias individuales en cómo la enseñanza facilita el aprendizaje de los estudiantes. Esta investigación fue conducida también con el propósito de aportar evidencias de validez de constructo para el cuestionario de evaluación estudiantil de la calidad en la enseñanza (cuestionario SET) y para el cuestionario de evaluación socio-cognitiva del proceso de aprendizaje (cuestionario SC).

La hipótesis de investigación es consistente con resultados obtenidos en varios trabajos previos (R. Ames y Lau, 1979; Bandura y cols., 1996; Grimes y cols., 2004; Svanum y Aigner, 2011; Wigfield y cols., 1997; Zimmerman, 2000). Estos trabajos en general reportan diferencias entre grupos de estudiantes en los puntajes obtenidos con cuestionarios SET, que se diferencian por alguna variable específica, por ejemplo, locus de causalidad, locus de control o tipos de contenidos de metas.

En general, las diferencias observadas no fueron relacionadas explícitamente, y en forma precisa, con el proceso de respuesta de los estudiantes al evaluar la enseñanza. Una excepción, respecto a esta observación, es el trabajo de R. Ames y Lau (1979) que consideró el proceso de respuesta de los estudiantes, pero, restringido solo al papel del locus de causalidad.

Algunas de estas investigaciones presentan limitaciones relacionadas con el cuestionario SET empleado. Adicionalmente, en general estas investigaciones tienen como unidad de análisis grupos o clases de estudiantes. Por tanto, la información que aportan no puede ser relacionada en forma directa con el proceso de respuesta a nivel de cada estudiante.

La presente investigación fue diseñada para superar varias de las limitaciones señaladas. Adicionalmente, la hipótesis central de esta investigación representa una posible explicación general de varios hallazgos previos. A su vez, permite establecer un vínculo explícito y preciso entre algunos hallazgos previos y el proceso de respuesta a cuestionarios SET. Además, en esta investigación la evaluación de la hipótesis propuesta contribuyó a entender mejor si las respuestas de los estudiantes a los cuestionarios SET, en general, aportan información fidedigna y útil sobre la enseñanza, o bien, presentan sesgos o inconsistencias, que limitan la utilidad de los datos recabados con estos instrumentos.

En los capítulos previos de esta tesis se aportaron diversas evidencias que apoyan claramente la hipótesis de investigación propuesta. En primer lugar, consideremos las evidencias obtenidas a partir del análisis de las respuestas de los estudiantes a los cuestionarios SET, con el modelo de múltiples facetas de Rasch (MFRM).

El patrón de interacciones entre elementos de las facetas en el modelo MFRM, empleado para modelar las respuestas a los cuestionarios SET resultaron consistentes con predicciones realizadas a partir de la hipótesis de investigación (véase capítulo 6). En primer lugar, se encontraron interacciones estadísticamente significativas, entre los elementos de la faceta estudiantes y faceta docentes, que además explicaron una proporción de la varianza con significación práctica (J. Cohen, 1992; Ferguson, 2009), para las nueve dimensiones de la enseñanza consideradas. También, las interacciones entre estudiantes e ítems encontradas fueron consistentemente interpretadas aplicando la hipótesis de investigación.

El análisis de las respuestas de los estudiantes al cuestionario SC con el modelo MFRM también aportó evidencias que respaldan la hipótesis de investigación (véase capítulo 7). La amplitud de la variabilidad en las medidas de los estudiantes y docentes, obtenidas a partir del análisis del cuestionario SC, tiene una relación con la amplitud de la variabilidad encontrada en las medidas de severidad de los estudiantes, que es consistente con la hipótesis de

investigación.

Así, si una parte substancial de la variabilidad observada en las severidades de los estudiantes es explicada por las variables socio-cognitivas consideradas, la gran variación observada en las severidades de los estudiantes no sería consistente con una baja variabilidad en las medidas de los estudiantes y en los efectos facilitadores de los docentes, obtenidos mediante el cuestionario SC. Sin embargo, la variabilidad relativamente elevada de las medidas encontradas de docentes y estudiantes, es consistente con los niveles observados de variabilidad en las severidades de los estudiantes. Por tanto, esto constituye una evidencia indirecta que apoya la hipótesis central de esta investigación.

Adicionalmente, se aportaron diversas evidencias que apoyan la hipótesis de investigación, a partir de modelos explicativos de la severidad global de los estudiantes, y del análisis de la relación entre las medidas de severidad y las medidas de las variables socio-cognitivas. En primer lugar, el modelado de las respuestas a los cuestionarios SET, con un modelo de MFRM con restricciones lineales en los parámetros de severidad (LLT-MFRM), aportaron evidencias que respaldan la hipótesis de investigación propuesta.

El modelo LLT-MFRM permite modelar explícitamente el proceso de respuesta de los estudiantes definiendo componentes o factores de severidad relacionados con las variables socio-cognitivas. A partir de las hipótesis de investigación se espera que las medidas de los componentes o factores de severidad sean distintos de cero y que el ajuste global entre los datos y el modelo sea aceptable. Además, se espera un ajuste aceptable para los componentes de la faceta de factores de severidad. Estas predicciones fueron consistentes con los resultados obtenidos.

En segundo lugar, los resultados del análisis del patrón de correlaciones entre las medidas de severidad y las medidas de las variables socio-cognitivas, también fueron consistentes con la hipótesis central de esta investigación. Estos resultados muestran claramente que las medidas de severidad presentaron una correlación significativamente distinta de cero con las medidas de las variables socio-cognitivas de los estudiantes.

En tercer lugar, el modelado de la severidad global mediante regresión robusta aportó evidencias adicionales que respaldan la hipótesis de investigación. Los efectos de las variables socio-cognitivas en los modelos de regresión, relacionados con los modelos teóricos del proceso de respuesta, propuestos en el capítulo 4, fueron en general estadísticamente significativos.

En la Figura 9.1 se representa la relación entre las variables socio-cognitivas y el proceso de respuesta de los estudiantes de acuerdo a las evidencias encontradas. Observar que la relación entre la severidad global y las variables socio-cognitivas está vinculada con la experiencia previa del estudiante y un plazo más largo que la duración de cierto curso. Por el contrario, la severidad local está relacionada con la experiencia actual de un estudiante en un curso.

En el capítulo 6 ambos tipos de severidad fueron consideradas y medidas. Por otra parte, en el capítulo 8 los análisis realizados se concentraron en la relación de la severidad global con las variables socio-cognitivas. La severidad global y local están relacionadas entre sí. En el capítulo 6 la severidad local se midió como la suma de la severidad global más la interacción entre un estudiante y el docente que evaluó.

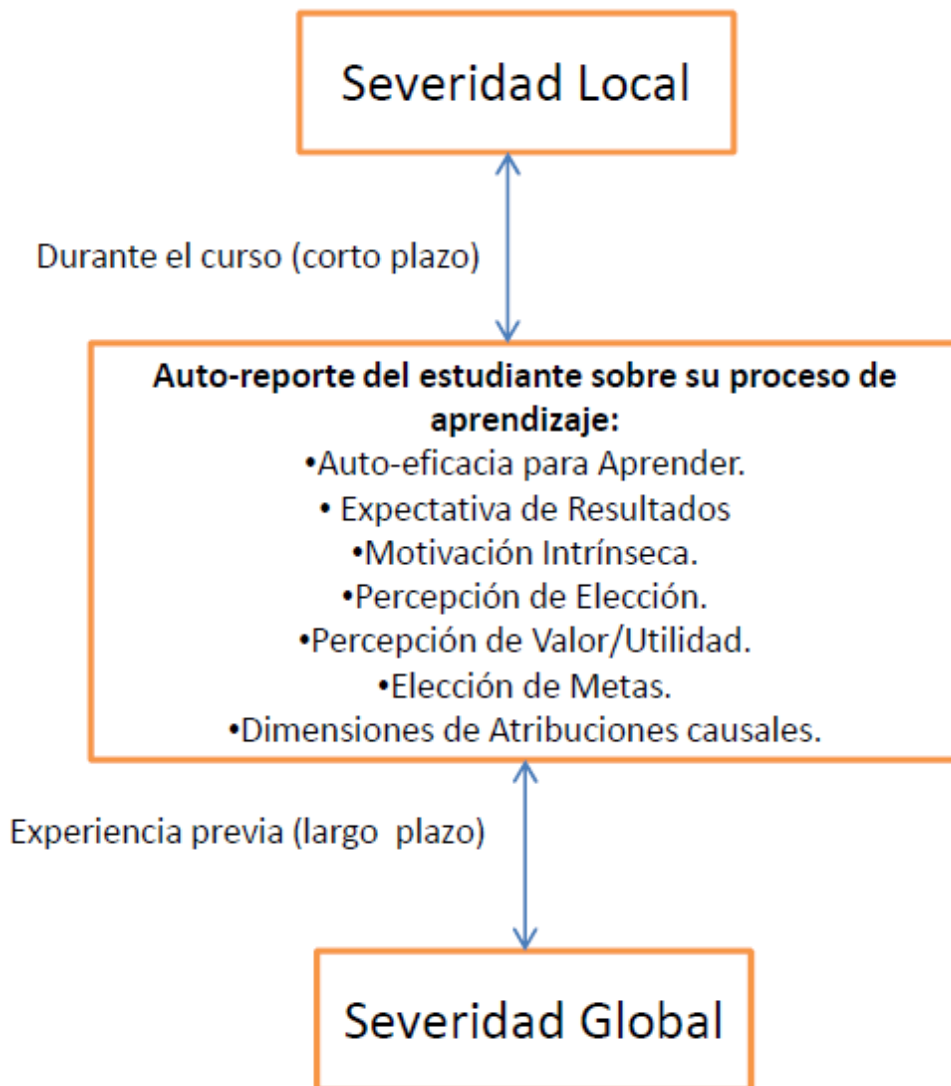


Figura 9.1: Variables socio-cognitivas y Proceso de Respuesta.

En la Figura 9.2 se representa un resumen conceptual de las relaciones que podrían existir entre el efecto facilitador de la enseñanza, la severidad global y local. En el capítulo 7 se midió el efecto facilitador para cada una de las variables consideradas (auto-eficacia para aprender, expectativas positivas de resultados, motivación intrínseca, percepción de elección, percepción de valor y utilidad, elección de metas, locus de causalidad, estabilidad y controlabilidad).

Un resultado básico que se obtuvo en el capítulo 7 fue que se encontró una gran variabilidad en el efecto facilitador entre los docentes. Adicionalmente, se encontraron diferencias en el efecto facilitador de los docentes para diferentes

estudiantes. Esto es consistente con las relaciones representadas en la Figura 9.2 entre las características de la enseñanza, de los estudiantes y las variables contextuales.

Por otra parte, la relación representada en la Figura 9.2 entre el efecto facilitador de la enseñanza y la severidad global, es consistente con los resultados obtenidos en el capítulo 8. A su vez, los resultados vinculados al análisis de interacciones entre la faceta estudiantes y docentes en el capítulo 6 es coherente con las relaciones representadas en la Figura 9.2 de las interacciones con el efecto facilitador de la enseñanza y con la severidad local.

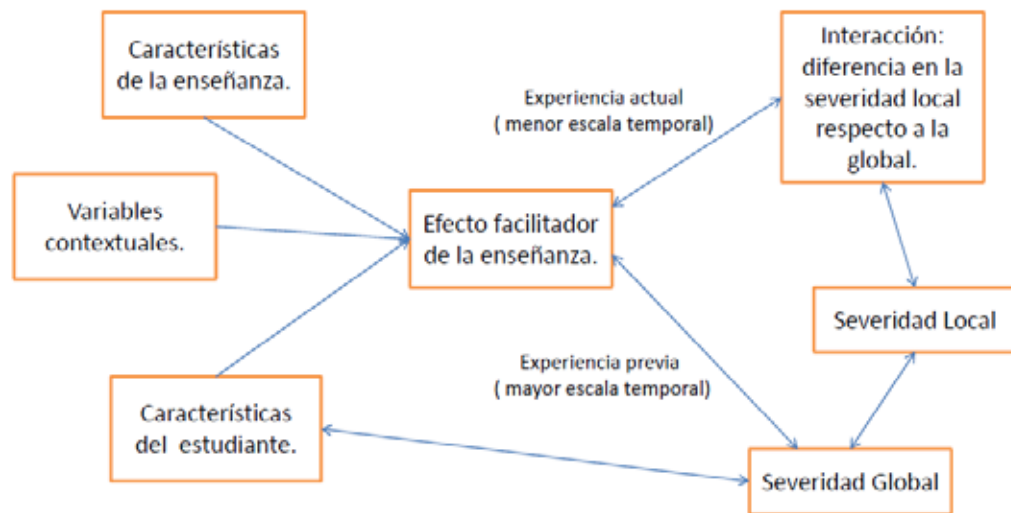


Figura 9.2: Relación entre severidad local, global y efecto facilitador.

9.2. Validez de constructo del cuestionario SET.

En los capítulos 6 y 8 se aportaron diversas evidencias que respaldan la validez de constructo del cuestionario SET. En primer lugar, los resultados encontrados son consistentes con una adecuada interpretación de los ítems. Los ítems de las sub-escalas en general muestran un elevado ajuste al modelo MFRM sin interacciones.

Adicionalmente, los valores encontrados de correlación punto-medida (r_{pm}) mostraron un nivel alto de consistencia entre las medidas de los ítems y el constructo medido por cada sub-escala. Además, los niveles de confiabilidad

de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) indicaron que los estudiantes son capaces de apreciar las diferencias que existen de contenido entre los ítems. Esto implica que es poco probable que exista un efecto halo a nivel de grupo (Eckes, 2011; Engelhard Jr, 2013; Myford y Wolfe, 2004). Por tanto, en general los estudiantes fueron capaces de distinguir entre aspectos conceptualmente distintos que se presenten en los contenidos de los ítems.

En segundo lugar, los resultados obtenidos mostraron que los estudiantes, como grupo, fueron capaces de distinguir entre los docentes según la magnitud de la calidad en la enseñanza en las dimensiones medidas por las sub-escalas estudiadas. Los docentes presentaron un buen nivel de ajuste al modelo para todas las sub-escalas consideradas. Adicionalmente, la medida de cada docente fue altamente consistente con el constructo medido por la escala, lo que se apreció al considerar los valores obtenidos de r_{pm} . Además, los niveles de confiabilidad de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) sugirió que los estudiantes fueron capaces de apreciar las diferencias que existen de calidad de la enseñanza entre los docentes. Por tanto, es poco probable que exista un efecto de restricción de rango a nivel de grupo (Engelhard Jr, 2013; Myford y Wolfe, 2004).

Por último, los resultados obtenidos en esta investigación mostraron que existen entre 4 y 5 estratos estadísticamente discernibles de severidad de los estudiantes. Esto mostró la existencia de diferencias importantes en cómo los estudiantes evalúan a los docentes (alto grado de subjetividad). Esto se reflejó también en el amplio rango de variación de las severidades.

Si bien los estudiantes presentaron diferencias importantes en su severidad al evaluar la enseñanza, varios resultados indicados previamente son consistentes con una aceptable capacidad para evaluar la enseñanza. Por otra parte, en el capítulo 8 se aportaron evidencias adicionales, que respaldan claramente que los estudiantes tienen habilidad razonable como evaluadores de la enseñanza como para aportar información útil.

Así, en el capítulo 8 se consideraron cuatro modelos teóricos que fueron propuestos y descritos en el capítulo 4. Estos modelos representan especificaciones o restricciones de la hipótesis central de esta investigación, que definen “hipótesis de validez” (R. Ames y Lau, 1979) de la evaluación estudiantil de la enseñanza. Un buen acuerdo con los modelos teóricos considerados, implicaría que la información obtenida a partir de las evaluaciones estudiantiles de la

enseñanza y proceso de aprendizaje aportan información útil y fidedigna. Esto fue fuertemente respaldado por los resultados obtenidos.

9.3. Validez de constructo del cuestionario SC.

En los capítulos 7 y 8 se aportaron diversas evidencias que respaldan la validez de constructo del cuestionario SC. En el capítulo 4 se discutieron en detalle los modelos teóricos que fueron la base para el análisis de validez de constructo de estas escalas. A continuación se resumen algunos de los aspectos más relevantes.

En primer lugar, los resultados de ajuste entre los datos y el modelo a nivel global y a nivel de cada faceta resultaron consistentes con un acuerdo razonable de las mediciones obtenidas con los criterios ideales y bastante exigentes implícitos en los modelo de múltiples facetas de Rasch. En relación al ajuste de las facetas, se encontró un buen ajuste en general. En particular el ajuste de los ítems fue muy elevado.

Además, los valores encontrados de r_{pm} fueron coherentes con un nivel elevado de consistencia entre las medidas de los ítems y el constructo medido por cada sub-escala. Adicionalmente, los niveles de confiabilidad de separación y los resultados encontrados con la prueba de efectos fijos (todos iguales) sugieren que los estudiantes son capaces de apreciar las diferencias que existen de contenido entre los ítems.

Por otra parte, en general los valores de dificultad de los ítems mostraron una distribución razonable a lo largo de la amplitud de variación de los constructos considerados. En general, las medidas de dificultad de los ítems resultaron consistentes con expectativas teóricas que surgen de análisis del significado del constructo y de los enunciados de los ítems.

Finalmente, los resultados ya discutidos sobre la evaluación de la hipótesis central de esta investigación y los resultados obtenidos en el capítulo 8 con modelos explicativos de la severidad global también representan una evidencia importante de la validez de constructo del cuestionario SC.

En síntesis, las evidencias aportadas en los capítulos 7 y 8 de esta tesis ofrecieron un amplio respaldo a la validez de constructo de este cuestionario.

9.4. Modelos teóricos del proceso de respuesta e Hipótesis de Validez.

La hipótesis central de esta investigación y la validez de constructo relacionada con los cuestionarios, explícitamente se analizó desde la perspectiva de cuatro modelos teóricos: el modelo SDT, el modelo TAS, el modelo Atribucional y el modelo de Auto-teorías. Estos modelos fueron propuestos y descritos en el capítulo 4 de esta tesis. A partir de la hipótesis central de esta investigación y de los cuatro modelos teóricos se derivaron una serie de predicciones teóricas, que fueron detalladamente contrastadas con los patrones observados en los datos.

Los modelos propuestos representan especificaciones o restricciones de la hipótesis central de esta investigación, que definen “hipótesis de validez” de la evaluación estudiantil de la enseñanza. Un buen acuerdo con los modelos teóricos considerados, significan que la información obtenida a partir de las evaluaciones estudiantiles de la enseñanza y proceso de aprendizaje aportan información válida. Así, a pesar de que existe un elevado efecto del evaluador, lo que implica un alto grado de subjetividad, no obstante, es posible obtener información fidedigna y útil. Las predicciones realizadas con los cuatro modelos teóricos fueron ampliamente respaldadas por las evidencias obtenidas en esta tesis.

En la Figura 9.3 se muestra una síntesis de los aspectos comunes de los modelos teóricos propuestos. Dado el grado de acuerdo entre las predicciones teóricas derivadas de los modelos y los patrones observados en los análisis realizados, la Figura 9.3 describe además las relaciones encontradas efectivamente entre el efecto facilitador de la enseñanza, la severidad global y la severidad local.

Las relaciones representadas en la Figura 9.3 tienen una implicación práctica importante. A pesar de que se encontraron evidencias claras que indican variaciones importantes en la severidad de los estudiantes, y por tanto, que las evaluaciones tienen un alto grado de subjetividad, no obstante, las diferencias en severidad están relacionadas con diferencias en el efecto facilitador de la enseñanza. Por tanto, reflejan diferencias en la calidad del proceso de aprendizaje y el efecto facilitador de la enseñanza. Es decir, representan variabilidad relevante para el constructo (Marsh y Roche, 1997; Messick, 1995).

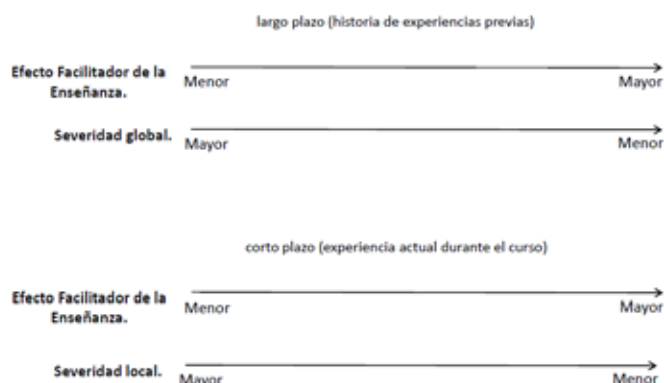


Figura 9.3: Hipótesis de Validez. Evaluación subjetiva, pero, fidedigna e informativa; refleja diferencias en la calidad del proceso de aprendizaje y el efecto facilitador de la enseñanza.

9.5. Aplicaciones prácticas.

Los aportes teóricos y metodológicos presentados en esta investigación pueden contribuir a un mejor uso e interpretación de instrumentos de evaluación de la calidad de la enseñanza y del proceso de aprendizaje. En forma más general, constituyen una contribución a una mejora práctica de la evaluación educativa. Muchas de las limitaciones presentes actualmente en la práctica de evaluación educativa podrían superarse considerablemente con la aplicación del marco teórico y metodológico aportado en esta tesis.

Adicionalmente, el presente trabajo muestra la factibilidad de una aproximación complementaria en la evaluación de la calidad en la enseñanza de los docentes. Esta aproximación incluye la medición y evaluación diagnóstica tanto del proceso de aprendizaje de los estudiantes como del efecto facilitador de los docentes.

Referencias

- Albright, L., y Malloy, T. E. (2000). Experimental validity: Brunswik, Campbell, Cronbach, and enduring issues. *Review of general psychology*, 4(4), 337.
- Albright, L., y Malloy, T. E. (2001). Brunswik's theoretical and methodological contributions to research in interpersonal perception. *The essential Brunswik: Beginnings, explications, applications*, 328–332.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, 84(3), 261.
- Ames, C., y Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of educational psychology*, 80(3), 260.
- Ames, R., y Lau, S. (1979). An attributional approach to the validity of student ratings of instruction. *Contemporary Educational Psychology*, 4(1), 26–39.
- Ames, R., y Lau, S. (1982). An attributional analysis of student help-seeking in academic settings. *Journal of Educational Psychology*, 74(3), 414.
- Anderson, C. A. (1983). Motivational and performance deficits in interpersonal settings: The effect of attributional style. *Journal of Personality and Social Psychology*, 45(5), 1136.
- Andrews, G. R., y Debus, R. L. (1978). Persistence and the causal perception of failure: Modifying cognitive attributions. *Journal of educational psychology*, 70(2), 154.
- Andrews, J. (2004). Teaching effectiveness. *Encyclopedia of applied psychology*, 3, 535–539.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. Descargado de <http://dx.doi.org/10.1007/BF02293814> doi: 10.1007/BF02293814
- Andrich, D. (1988). *Rasch models for measurement: Sage publications*

- (Vol. 68). Sage Publications.
- Arkin, R. M., y Maruyama, G. M. (1979). Attribution, affect, and college exam performance. *Journal of Educational Psychology*, 71(1), 85–93.
- Arreola, R. A. (2006). *Develong a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems* (Vol. 86). Jossey-Bass.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review*, 64(6p1), 359.
- Babcock, B., y Albano, A. D. (2012). Rasch scale stability in the presence of item parameter and trait drift. *Applied Psychological Measurement*, 36(7), 565–580.
- Baker, F. B., y Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. CRC Press.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.
- Bandura, A. (1978a). Reflections on self-efficacy. *Advances in behaviour research and therapy*, 1(4), 237–269.
- Bandura, A. (1978b). The self system in reciprocal determinism. *American psychologist*, 33(4), 344–358.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, 37(2), 122.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental psychology*, 25(5), 729.
- Bandura, A. (1990). Self-regulation of motivation through anticipatory and self-reactive mechanisms. En *Nebraska symposium on motivation. nebraska symposium on motivation* (Vol. 38, p. 69).
- Bandura, A. (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. En *Perspectives on motivation: Nebraska symposium on motivation* (Vol. 38, pp. 69–164).
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist*, 28(2), 117–148.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1–26.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents*, 5(307-337).
- Bandura, A., Barbaranelli, C., Caprara, G. V., y Pastorelli, C. (1996). Mul-

- tifaceted impact of self-efficacy beliefs on academic functioning. *Child development*, 1206–1222.
- Bandura, A., y Walters, R. H. (1977). Social learning theory.
- Bandura, A., y Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of personality and social psychology*, 56(5), 805.
- Baumeister, R. F., y Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3), 497.
- Bernstein, I. H., y Nunnally, J. (1994). Psychometric theory. *New York: McGraw-Hill*.
- Berry, J. M. (1999). Memory self-efficacy in its social cognitive context.
- Berry, J. M., West, R. L., y Dennehey, D. M. (1989). Reliability and validity of the memory self-efficacy questionnaire. *Developmental Psychology*, 25(5), 701.
- Bianchini, S., Lissoni, F., y Pezzoni, M. (2013). Instructor characteristics and students' evaluation of teaching effectiveness: evidence from an italian engineering school. *European Journal of Engineering Education*, 38(1), 38–57.
- Black, A. E., y Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science education*, 84(6), 740–756.
- Bond, T., y Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences*. Routledge.
- Braskamp, L. A., y Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance. jossey-bass higher and adult education series*. ERIC.
- Brickman, P., Rabinowitz, V. C., Karuza, J., Coates, D., Cohn, E., y Kidder, L. (1982). Models of helping and coping. *American psychologist*, 37(4), 368–384.
- Brown, S. (2011). Bringing about positive change in the higher education student experience: a case study. *Quality Assurance in Education*, 19(3), 195–207.
- Brunswik, E. (1947). Systematic and representative design of psychological experiments. En *Proceedings of the berkeley symposium on mathematical*

statistics and probability.

- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3), 193.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.
- Burry, J. A., y Shaw, D. (1988). Defining teacher effectiveness on a continuum: A rasch model approach.
- Campbell, D. T., y Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Centra, J. A. (1976). The influence of different directions on student ratings of instruction. *Journal of Educational Measurement*, 13(4), 277–282.
- Chang, T.-S. (2000). An application of regression models with student ratings in determining course effectiveness.
- Chernick, M. R. (2008). *Bootstrap methods: a guide for practitioners and researchers*. Hoboken. Nj: Wiley.
- Cohen, E. H. (2005). Student evaluations of course and teacher: Factor analysis and ssa approaches. *Assessment & Evaluation in Higher Education*, 30(2), 123–136.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Coronado, M. (2009). *Competencias docentes: Ampliación, enriquecimiento y consolidación de la práctica profesional*. Noveduc.
- Cronbach, L. J. (1988). Five perspectives on validity argument. *Test validity*, 3–17.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1), 105.
- Deci, E. L. (1972). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational behavior and human performance*, 8(2), 217–229.
- Deci, E. L. (1975a). Intrinsic motivation. *Springer Sciences*.
- Deci, E. L. (1975b). Notes on the theory and metatheory of intrinsic motivation. *Organizational behavior and human performance*, 15(1), 130–145.
- Deci, E. L., y Cascio, W. F. (1972). Changes in intrinsic motivation as a function of negative feedback and threats.

- Deci, E. L., Koestner, R., y Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of educational research*, 71(1), 1–27.
- Deci, E. L., y cols. (1973). Sex differences, positive feedback and intrinsic motivation.
- Deci, E. L., y Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. *Advances in experimental social psychology*, 13, 39–80.
- Deci, E. L., y Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. *Springer*.
- Deci, E. L., y Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227–268.
- Deci, E. L., y Ryan, R. M. (2002). Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research*, 3–33.
- Deci, E. L., y Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3), 182.
- Deci, E. L., Ryan, R. M., y Williams, G. C. (1996). Need satisfaction and the self-regulation of learning. *Learning and individual differences*, 8(3), 165–183.
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., y Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational psychologist*, 26(3-4), 325–346.
- de Jong, J., y Linacre, J. (1993). Estimation methods, statistical independence and global fit. *Rasch Measurement Transactions*, 7(2), 296–297.
- Draper, N. R., y Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, 41(10), 1040.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.
- Dweck, C. S., y Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological review*, 95(2), 256.
- Dwinell, P. L., y Higbee, J. L. (1993). Students’ perceptions of the value of

- teaching evaluations. *Perceptual and Motor Skills*, 76(3), 995–1000.
- Eckes, T. (2009). Many-facet rasch measurement. *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment*.
- Eckes, T. (2011). Introduction to many-facet rasch measurement. *Frankfurt: Peter Lang*.
- Edgington, E., y Onghena, P. (2007). *Randomization tests*. CRC Press.
- Efron, B. (1978). Regression and anova with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, 73(361), 113–121.
- Efron, B., y Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Ellis, J. L., y Van den Wollenberg, A. L. (1993). Local homogeneity in latent trait models. a characterization of the homogeneous monotone irt model. *Psychometrika*, 58(3), 417–429.
- Embretson, S. (1994). Applications of cognitive design systems to test development. En *Cognitive assessment* (pp. 107–135). Springer.
- Embretson, S., y Reise, S. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Engelhard Jr, G. (1996). Constructing rater and task banks for performance assessments. *Journal of outcome measurement*, 1(1), 19–33.
- Engelhard Jr, G. (2013). *Invariant measurement: Using rasch models in the social, behavioral, and health sciences*. Routledge.
- Everitt, B. (1984). *An introduction to latent variable models*. Chapman and Hall.
- Farrokhi, F., y Esfandiari, R. (2011). A many-facet rasch model to detect halo effect in three types of raters. *Theory and Practice in Language Studies*, 1(11), 1531–1540.
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10(2), 149–172.
- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532.
- Fischer, G. H. (1977). Linear logistic trait models: Theory and application. *Structural models of thinking and learning*, 203–225.
- French, G. A., y Dodd, B. G. (1998). Parameter recovery for the rating scale

- model using parscale. *Journal of outcome measurement*, 3(2), 176–199.
- Frey, P. W., Leonard, D. W., y Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal*, 12(4), 435–444.
- Frieze, I., y Weiner, B. (1971). Cue utilization and attributional judgments for success and failure. *Journal of Personality*, 39(4), 591–605.
- Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review*, 3(1), 5–39.
- Gredler, M. E. (2009). *Learning and instruction: Theory into practice*. Pearson.
- Green, K. E., y Smith, R. M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics*, 12(4), 369–381.
- Grimes, P. W., Millea, M. J., y Woodruff, T. W. (2004). Grades—who's to blame? student evaluation of teaching and locus of control. *The Journal of Economic Education*, 35(2), 129–147.
- Hammond, K. R., y Stewart, T. R. (2001). *The essential brunswik: Beginnings, explications, applications*. Oxford University Press.
- Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, 55(4), 577–601.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.
- Huber, P. (1981). J. 1981. robust statistics. *New York: John Wiley*.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, 112(3), 527.
- Kasser, T., y Ryan, R. M. (1996). Further examining the american dream: Differential correlates of intrinsic and extrinsic goals. *Personality and social psychology bulletin*, 22(3), 280–287.
- Kukla, A. (1972). Foundations of an attributional theory of performance. *Psychological Review*, 79(6), 454.
- Lepper, M., y Greene, D. (s.f.). Turning play into work: Effects of adult surveillance and motivation 93 extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*.
- Lepper, M. R., Greene, D., y Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and social Psy-*

- chology*, 28(1), 129.
- Lepper, M. R., Sethi, S., Dialdin, D., y Drake, M. (1997). Intrinsic and extrinsic motivation: A developmental perspective. *Developmental psychopathology: Perspectives on adjustment, risk, and disorder*, 23–50.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Linacre, J. M. (1989a). Objectivity for judge-intermediated certification examinations.
- Linacre, J. M. (1989b). Rasch models from objectivity: A generalization.
- Linacre, J. M. (1994). *Many-facet rasch measurement*. Mesa Press.
- Linacre, J. M. (2005). Fit to models: Rasch model vs. correlation model. *Rasch measurement transactions*, 19(3), 1029.
- Linacre, J. M. (2006). Data variance explained by rasch measures. *Rasch Measurement Transactions*, 20(1), 1045.
- Linacre, J. M. (2010). When to stop removing items and persons in rasch misfit analysis. *Rasch Measurement Transactions*, 23(4), 1241.
- Linacre, J. M. (2014). *A user's guide to facets rasch-model computer programs program manual 3.71.4 by*.
- Linacre, J. M. (2015). Facets: Facets computer program for many-facet rasch measurement [Manual de software informático]. Beaverton, Oregon. Descargado de <http://www.winsteps.com>
- Linacre, J. M., y cols. (2002). Optimizing rating scale category effectiveness. *J Appl Meas*, 3(1), 85–106.
- Lord, F., y cols. (1968). Novick. *MRStatistical theories of mental test scores*.
- Madichie, N. O. (2011). Students' evaluation of teaching (set) in higher education: A question of reliability and validity. *The Marketing Review*, 11(4), 381–391.
- Manly, B. F. (1991). Monte carlo and other computer-intensive methods. En *Randomization and monte carlo methods in biology* (pp. 21–30). Springer.
- Manly, B. F. (2006). *Randomization, bootstrap and monte carlo methods in biology* (Vol. 70). CRC Press.
- Marlin Jr, J. W. (1987). Student perceptions of end-of-course evaluations. *The Journal of Higher Education*, 704–716.
- Marsh, H. W. (1981). Students' evaluations of tertiary instruction: Testing the applicability of american surveys in an australian setting. *Australian*

Journal of Education, 25(2), 177–193.

- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International journal of educational research*, 11(3), 253–388.
- Marsh, H. W., y Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological bulletin*, 97(3), 562.
- Marsh, H. W., y Hocevar, D. (1991). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education*, 7(1), 9–18.
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., y Trautwein, U. (2009). Exploratory structural equation modeling, integrating cfa and efa: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476.
- Marsh, H. W., y Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American psychologist*, 52(11), 1187.
- Marsh, H. W., Tourón, J., y Wheeler, B. (1985). Students' evaluations of university instructors: The applicability of american instruments in a spanish setting. *Teaching and teacher Education*, 1(2), 123–138.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. Descargado de <http://dx.doi.org/10.1007/BF02296272> doi: 10.1007/BF02296272
- Masters, G. N., y Wright, B. D. (1997). The partial credit model. En *Handbook of modern item response theory* (pp. 101–121). Springer.
- McAuley, E., Duncan, T. E., y Russell, D. W. (1992). Measuring causal attributions: The revised causal dimension scale (cdsii). *Personality and Social Psychology Bulletin*, 18(5), 566–573.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Meyer, J. P. (1978). Dimensions of causal attribution for success and failure: A multivariate investigation.
- Meyer, J. P. (1980). Causal attribution for success and failure: A multivariate

- investigation of dimensionality, formation, and consequences. *Journal of Personality and Social Psychology*, 38(5), 704.
- Miller, G. A., Galanter, E., y Pribram, I. (1960). Plans and the structure of behavior. *New York: Holt, Rinehart and Winston*.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Myford, C. M., y Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet rasch measurement: Part ii. *Journal of applied measurement*, 5(2), 189–227.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological review*, 91(3), 328.
- Obiekwe, J. C. (1999). The multidimensional character of teaching effectiveness: A comparative analysis of student evaluation responses of full and part-time faculty.
- Paswan, A. K., y Young, J. A. (2002). Student evaluation of instructor: A nomological investigation using structural equation modeling. *Journal of Marketing Education*, 24(3), 193–202.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary educational psychology*, 25(1), 92–104.
- Pintrich, P. R., y cols. (1991). A manual for the use of the motivated strategies for learning questionnaire (mslq).
- Quezada, R. A. G. (2011). La responsabilidad social universitaria como desafío para la gestión estratégica de la educación superior: el caso de España. *Revista de educación*(355), 109–133.
- R Development Core Team. (2008). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <http://www.R-project.org> (ISBN 3-900051-07-0)
- Reddy, T. (2011). *Applied data analysis and modeling for energy engineers and scientists*. Springer Science & Business Media.
- Reddy, T., Saman, N., Claridge, D., Haberl, J., Turner, W., y Chalifoux, A. (1997). Baseline methodology for facility-level monthly energy use - part 1: theoretical aspects. En *Ashrae transactions* (pt 2 ed., Vol. 103, pp. 336–347). ASHRAE.
- Reeve, J. (2006). Teachers as facilitators: What autonomy-supportive teachers

- do and why their students benefit. *The Elementary School Journal*, 106(3), 225–236.
- Renaud, R. D., y Murray, H. G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education*, 46(8), 929–953.
- Reynolds, C., y Fels, M. (1988). Reliability criteria for weather adjustment of energy billing data. En *Proceedings of aceee* (pp. 10–237).
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1), 1.
- Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American psychologist*, 45(4), 489.
- Rupp, A. A., y Zumbo, B. D. (2004). A note on how to quantify and report whether irt parameter invariance holds: When pearson correlations are not enough. *Educational and Psychological Measurement*, 64(4), 588–599.
- Ryan, R. (2011). Intrinsic motivation inventory, self-determination theory questionnaire packet. *Self-Determination Theory Questionnaires website*. Descargado de <http://selfdeterminationtheory.org/intrinsic-motivation-inventory/>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3), 450.
- Ryan, R. M., y Connell, J. P. (1989). Perceived locus of causality and internalization: examining reasons for acting in two domains. *Journal of personality and social psychology*, 57(5), 749.
- Ryan, R. M., Connell, J. P., y Deci, E. L. (1985). A motivational analysis of self-determination and self-regulation in education. *Research on motivation in education: The classroom milieu*, 2, 13–51.
- Ryan, R. M., y Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54–67.
- Ryan, R. M., y Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68.
- Ryan, R. M., y Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. *Guilford*.

- Samejima, F. (1983). Some methods and approaches of estimating the operating characteristics of discrete item responses. *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord*, 159–182.
- Scheiblechner, H. (1972). Das lernen und losen komplexer denkaufgaben. *Zeitschrift für experimentelle und angewandte Psychologie*, 19, 476–506.
- Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational psychologist*, 25(1), 71–86.
- Schunk, D. H., Meece, J. R., y Pintrich, P. R. (2014). *Motivation in education: Theory, research, and applications*. Pearson Higher Ed.
- Seldin, P. (1989). How colleges evaluate professors. 1988 vs. 1983. *AAHE Bulletin*, 41(7), 3–7.
- Seldin, P. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*. Anker Publishing Company.
- Shell, D. F., Murphy, C. C., y Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, 81(1), 91.
- Shevlin, M., Banyard, P., Davies, M., y Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25(4), 397–405.
- Skinner, E. A., y Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, 85(4), 571.
- Smith, R., Linacre, J., y Smith, E. (2003). Guidelines for manuscripts. *Journal of Applied Measurement*, 4(2), 198–204.
- Smith, R. M., y cols. (1995). Using item mean squares to evaluate fit to the rasch model.
- Svanum, S., y Aigner, C. (2011). The influences of course effort, mastery and performance goals, grade expectancies, and earned course grades on student ratings of course satisfaction. *British Journal of Educational Psychology*, 81(4), 667–679.
- Tapp, S. R. (1984). Brunswik's lens model: A review and extension to consumer research. *NA-Advances in Consumer Research Volume 11*.
- Theall, M., Abrami, P. C., y Mets, L. A. (2001). The student ratings debate: Are they valid? how can we best use them?
- Trudeau, R. J. (2013). *Introduction to graph theory*. Courier Corporation.

- UDELAR. (2012). Documento de orientación para la carrera docente en la udelar. Descargado de [http://www.expe.edu.uy/expe/adjuntos.nsf/0/D6FAC6310E13E47D03257A2300699DA4/\\$File/Dist.%20393.12.PDF](http://www.expe.edu.uy/expe/adjuntos.nsf/0/D6FAC6310E13E47D03257A2300699DA4/$File/Dist.%20393.12.PDF)
- URUGUAY. (2008). Ley general de educación, ley 18.437. Descargado de http://www.mec.gub.uy/innovaportal/file/431/1/ley_de_educacion_dic2014.pdf
- Vallaes, F. (2009). "responsabilidad social universitaria": una nueva filosofía de gestión ética e intencionales para las universidades. *Educación superior y sociedad*, 13(2), 191–220.
- Vansteenkiste, M., Lens, W., y Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational psychologist*, 41(1), 19–31.
- Vansteenkiste, M., Niemiec, C. P., y Soenens, B. (2010). The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions. En *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 105–165). Emerald Group Publishing Limited.
- Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., y Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of personality and social psychology*, 87(2), 246.
- Venables, W. N., y Ripley, B. D. (2013). *Modern applied statistics with s-plus*. Springer Science & Business Media.
- Wainer, H., y Braun, H. I. (2013). *Test validity*. Routledge.
- Watkins, D. (1992). Evaluating the effectiveness of tertiary teaching: A hong kong perspective. *Educational Research Journal*, 7, 60–67.
- Watkins, D., y Akande, A. (1992). Student evaluations of teaching effectiveness: A nigerian investigation. *Higher Education*, 24(4), 453–463.
- Watkins, D., Marsh, H. W., y Young, D. (1987). Evaluating tertiary teaching: A new zealand perspective. *Teaching and Teacher Education*, 3(1), 41–53.
- Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. John Wiley & Sons.
- Weiner, B. (1974). An attributional interpretation of expectancy-value theory. *Cognitive views of human motivation*, 51–69.
- Weiner, B. (1983). Some methodological pitfalls in attributional research.

- Journal of Educational Psychology*, 75(4), 530.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), 548.
- Weiner, B. (1986). *An attribution theory of motivation and emotion*.
- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Sage.
- Weiner, B. (2013). *Human motivation*. Psychology Press.
- Weiner, B., Nierenberg, R., y Goldstein, M. (1976). Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy of success¹. *Journal of Personality*, 44(1), 52–68.
- Weiner, B., Russell, D., y Lerman, D. (1979). The cognition–emotion process in achievement-related contexts. *Journal of personality and social psychology*, 37(7), 1211.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbretton, A. J., Freedman-Doan, C., y Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of educational psychology*, 89(3), 451.
- Williams, G., Ryan, R., y Deci, E. (2011a). Health-care, self-determination theory questionnaire packet. *Self-Determination Theory Questionnaires website*.
- Williams, G., Ryan, R., y Deci, E. (2011b). Perceived competence scales, self-determination theory questionnaire packet. *Self-Determination Theory Questionnaires website*.
- Williams, G. C., y Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: a test of self-determination theory. *Journal of personality and social psychology*, 70(4), 767.
- Willmott, C. J., y Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1), 79–82.
- Wills, T. A., y DePaulo, B. M. (1991). Interpersonal analysis of the help-seeking process. *Handbook of social and clinical psychology*, 162, 350–375.
- Wilson, M., y De Boeck, P. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer New York, NY.
- Wilson, T. D., y Linville, P. W. (1982). Improving the academic performance

- of college freshmen: Attribution therapy revisited. *Journal of personality and social psychology*, 42(2), 367.
- Wood, R., y Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of personality and social psychology*, 56(3), 407.
- Wright, B. D., y Masters, G. N. (1990). Computation of outfit and infit statistics. *Rasch Measurement Transactions*, 3(4), 84–85.
- Wright, B. D., y Stone, M. H. (1999). Measurement essentials. *Wilmington. Wide Range Inc*, 221.
- Wright, R. E. (2006). Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal*, 40(2), 417.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. En *Self-efficacy in changing societies*.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, 25(1), 82–91.

Glosario

Amotivación es un estado en el cual la persona no está motivada a llevar a cabo un comportamiento, o bien, el comportamiento no está mediado por intencionalidad. 36

Atribución causal es una explicación causal realizada por un sujeto a un resultado personal. Es decir, las atribuciones causales son interpretaciones sobre las causas percibidas de un resultado personal. 19

Autoeficacia es el grado en que una persona confía en su habilidad para realizar una tarea. 17

Expectativa de resultado es la relación percibida entre realizar una tarea con éxito y obtener un resultado determinado por efecto de esa realización. 17

Metas extrínsecas refiere a metas cuyo valor es dependiente de evaluaciones de otros. 40

Metas intrínsecas intrínsecas son un tipo de metas inherentemente valiosas o satisfactoria para el individuo. 40

Motivación intrínseca es un tipo de motivación que no implica una recompensa externa aparente. Un comportamiento es intrínsecamente motivado cuando el motivo para realizarlo está relacionado con la respuesta afectiva que se origina por la realización del comportamiento en sí. Desde el punto de vista fenomenológico, la persona que lleva a cabo un comportamiento intrínsecamente motivado, experimenta disfrute, interés e inmersión en la tarea que realiza. 29

Necesidad de autonomía refiere a la necesidad del ser humano de percibir que tiene elecciones y control sobre sus propias acciones. 29

Necesidad de competencia es la necesidad del ser humano de percibir que es capaz de actuar en forma competente y eficaz. 28

Necesidad de relacionamiento refiere a la necesidad del ser humano de

percibir que está conectado a otros seres humanos a través de relaciones positivas. 29

Necesidad psicológica básica es una necesidad psicológica que se caracteriza por ser innata, universal, y constituir un nutriente psicológico básico para el crecimiento, desarrollo y bienestar psicológico. 28

Proyección refiere a la atribución de auto-aprobaciones y auto-reprobaciones a otros, imaginando que estos otros aprobaran o desaprobaran en función de ciertos comportamientos o resultados. La proyección es un mecanismo en el que en parte se basa la regulación introyectada. 37



APÉNDICES



Maestría en Enseñanza Universitaria

Comisión Sectorial de Enseñanza
Área Social y Artística
Consejo de Formación en Educación



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



comisión sectorial
de enseñanza



Facultad
de Humanidades
y Ciencias de la Educación



Consejo de
Formación en
Educación

Apéndice 1

Consentimiento Informado

Consentimiento Informado WEB Aplicado en el Sistema de Evaluación de la Enseñanza y Proceso de Aprendizaje de Facultad de Medicina:

El Departamento de Educación Médica es responsable de llevar a cabo las Evaluaciones Institucionales de la Enseñanza y Aprendizaje en Facultad de Medicina, para lo cual se recaban datos relacionados con la enseñanza y proceso de aprendizaje.

Los datos recabados serán empleados para mejorar la enseñanza, aprendizaje y rendimiento académico de los estudiantes. También en tareas de evaluación educativa. Los datos recabados se utilizarán para realizar diagnósticos de fortalezas y debilidades en la enseñanza y aprendizaje, y se utilizarán para informar la toma de decisiones necesarias para implementación de mejoras.

Los datos recabados incluirán datos relacionados con su opinión sobre la enseñanza y su proceso de aprendizaje. No incluirán datos personales sensibles.

Los datos recabados del proceso de aprendizaje de cada estudiante y de la enseñanza serán confidenciales. Estos datos están protegidos por secreto estadístico (Ley N° 16.616) y por la Ley de Datos Personales y Acción de Habeas Data (Ley N° 18.331).

Los estudiantes recibirán beneficios a través de la producción de insumos y mejoras en la enseñanza y aprendizaje. El único costo es el tiempo de respuesta que es menor a 15 minutos en promedio. Los docentes recibirán insumos para su autoevaluación y la mejora de su práctica de enseñanza.

Los datos se emplearán en tareas de investigación para estudiar los procesos de enseñanza y aprendizaje y su relación, en el contexto educativo en el cual se



desarrollan. También se utilizarán para investigar la validez de los instrumentos de medición (formularios de evaluación de la enseñanza y el aprendizaje) y procedimientos de evaluación.

Estas tareas de investigación son necesarias para poder llevar a cabo correctamente las tareas de evaluación educativa y de producción de insumos para promover una mejor enseñanza y aprendizaje. Es mediante estas tareas de investigación que se pueden elaborar formularios y procedimientos de evaluación válidos.

En caso de preguntas o dudas escribir un mail al responsable del Sistema de Evaluación de la Enseñanza y Aprendizaje en Facultad de Medicina - Prof. Adj. Roberto Cáceres (mail: rcaceres@fmed.edu.uy).

Si está de acuerdo con que los datos aportados se utilicen en futuras investigaciones con tales fines indíquelo a continuación:

| | Estoy de acuerdo: | No estoy de acuerdo: |
|--------------------------------------------------------------------------------------------|-----------------------|-----------------------|
| Si ha leído y entendido la información previa indíquelo si está de acuerdo a continuación: | <input type="radio"/> | <input type="radio"/> |

Apéndice 2

Código empleado en la anonimización de datos.

A continuación se muestra el código de dos funciones básicas en la implementación de la funcionalidad de la aplicación desarrollada para la anonimización de los datos.

```
#####
### function anonimo
### Programador: R. D. Cáceres-Bauer
### R version 3.0.3 (2014-03-06) -- "Warm Puppy"
#####

anonimo<-function(data){
  L<-nrow(data);
  ru_vector<-runif(L);
  r_ord<-order(ru_vector);
  IDoldNewOrd<-data$IDold[r_ord]
  idElem.df<-idElemento(IDoldNewOrd);
  dataNewOrd<-data[r_ord,];
  dataNewOrd$IDold<-idElem.df$idElem;
  return(dataNewOrd)
}#End anonimo

#####
```

```
#####
### function idElemento(IDold)
### Programador: R. D. Cáceres-Bauer
### R version 3.0.3 (2014-03-06) -- "Warm Puppy"
#####

idElemento <-function(IDold){
  L=length(IDold)
  idElem <-rep(0,L);
  seq<-1:L;
  seqdif<-seq;
  IDoldC<-IDold;
  i<-0;
  while(length(seqdif)>0){
    i<-i+1;
    index<- IDold == IDoldC[1]
    idElem[index]<-i;
    seqdif<-setdiff(seqdif, seq[index])
    IDoldC <- IDold[seqdif];
  }#End while
  return(idElem.df)
}# End idEst
#####
```