

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

EVALUACIÓN Y ACREDITACIÓN UNIVERSITARIA.

Roberto Daniel Cáceres Bauer^{1,2}

1 Departamento de Educación Médica, Facultad de Medicina, Universidad de la República, Uruguay, rcaceresb@gmail.com

2 Unidad de Estadística, Escuela de Nutrición, Área Investigación, Universidad de la República, Uruguay.

RESUMEN

Un primer objetivo de este trabajo es evaluar la validez y propiedades psicométricas de un cuestionario de evaluación de la calidad en la enseñanza. Una hipótesis en esta 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 promueve el aprendizaje de los estudiantes. Un segundo objetivo es evaluar la validez de algunas predicciones teóricas derivadas de esta hipótesis. La principal predicción realizada es que la faceta estudiantes debería tener un menor ajuste al modelo considerado 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. El cuestionario considerado consiste en 49 ítems organizados en 9 sub-escalas. Estas sub-escalas fueron desarrolladas siguiendo procedimientos requeridos para obtener instrumentos con una adecuada validez (contenido y aparente). La muestra de calibración correspondió a 3243 estudiantes que evaluaron en conjunto 98 docentes de 10 actividades de enseñanza y siete disciplinas diferentes y corresponden a estudiantes que respondieron las sub-

escalas 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 en el año 2015. Las respuestas a cada una de las sub-escalas fueron analizadas por separado mediante un modelo de múltiples facetas de Rasch. 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.

Palabras clave: Evaluación estudiantil de la enseñanza, calidad de la enseñanza, modelos de múltiples facetas de Rasch.

1. INTRODUCCIÓN

Los cuestionarios de evaluación estudiantil de la enseñanza son instrumentos utilizados en universidades y otras instituciones de educación superior. Varios ejemplos de estos instrumentos consideran múltiples dimensiones, y están organizadas en sub-escalas 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. Algunas escalas de ítems 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 y cols., 2001). Un correcto uso e interpretación de estos instrumentos en la práctica educativa requiere una buena comprensión de la validez y significado de las medidas obtenidas con estos instrumentos.

2. PROBLEMÁTICA PROPUESTA Y CONTEXTO

A pesar de la importancia práctica que tienen los cuestionarios de evaluación estudiantil de la calidad de la enseñanza, y de la gran cantidad de trabajos de investigaciones sobre estos instrumentos, existen algunos vacíos de conocimiento con relación a su validez. Si bien se han realizado algunos estudios sobre las actitudes de los estudiantes acerca de estos instrumentos (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 SETs (Theall y cols., 2001). Algunas preguntas relacionadas con la presente investigación son: ¿Qué tan buena es la interpretación de los estudiantes de los criterios de evaluación contenidos en los ítems?; ¿Los estudiantes evalúan a los docentes en forma consistente?; ¿Existen muchas diferencias entre las formas de evaluar de los estudiantes?; ¿La severidad del estudiante al evaluar a un docente interactúa de alguna forma con las características de la enseñanza que está siendo evaluada?

3. OBJETIVO GENERAL

Un primer objetivo de este trabajo es evaluar la validez y propiedades psicométricas de un cuestionario de evaluación de la calidad en la enseñanza. Una hipótesis fundamental en esta 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 promueve o facilita el aprendizaje de los estudiantes. Un segundo objetivo es evaluar la validez de algunas predicciones teóricas derivadas de esta hipótesis de investigación. La principal predicción realizada 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.

4. METODOLOGÍA

El cuestionario considerado consiste 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 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). La participación de los estudiantes fue voluntaria, previo consentimiento informado y los datos fueron anonimizados. Esta investigación fue evaluada y aprobada por el Comité de Ética de Facultad de Medicina de la Universidad de la República.

5. RESULTADOS

5.1. Ajuste global entre los datos y el modelo.

Se encontró un buen ajuste global entre el modelo y los datos para todas las sub-escalas consideradas (Tabla 1). MAE y RMSE son dos métricas utilizadas para evaluar el ajuste absoluto, que muestran un mejor ajuste entre los datos y el modelo, cuanto más pequeño es su valor (cerca no a cero). Los valores encontrados son relativamente pequeños como muestra el valor normalizado, CV(RMSE), que sugiere que el ajuste absoluto para todas las escalas consideradas es aceptable. Adicionalmente, el modelo considerado para cada una de las escalas explica un porcentaje elevado de la varianza observada, como muestra los valores del índice de bondad de ajuste (GFI), que describe la proporción de la varianza observada que es explicada por el modelo. Por otra parte, los valores de InfitMS y OutfitMS para el ajuste global del modelo, sugieren un buen ajuste. En forma similar la media y desvío estándar para los residuos estandarizados para cada escala tienen valores muy cercanos a los esperados en caso de un perfecto ajuste entre los datos y el modelo considerado, lo que también sugiere un buen ajuste global.

Tabla 1 Ajuste Global.

	VALR E1	INTR E2	RELP E3	TRDF E4	CLOR E5	ENTC E6	APPR E7	TRPR E8	EVAL E9
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
GFI	0,59	0,56	0,64	0,57	0,62	0,58	0,58	0,66	0,53
InfitMS	0,99	1,00	0,99	1,00	0,99	0,99	0,99	0,99	1,00
OutfitMS	0,98	0,98	0,97	0,98	0,97	0,97	0,98	0,94	0,99
Media StRes	0,00	0,00	0,02	0,00	0,01	0,01	0,00	0,01	0,00
SD StRes	0,98	0,98	0,94	0,99	0,96	0,96	0,97	0,95	0,99

5.2. Resultados del 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 2,3 y 4). Los resultados muestran también una excelente confiabilidad de separación (con confiabilidad de Rasch) para cada una de las dimensiones de la calidad en la enseñanza consideradas y para la dificultad de los ítems (Tabla 5). En relación al ajuste de la faceta estudiantes se encontró un nivel de ajuste menor que para las otras facetas consideradas. 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.

Tabla 2 Faceta Ítems

Items	Medida	S.E.	InfitMS	OutfitMS	PtMea
VALR1	0,24	0,02	1,04	1,04	0,77
VALR2	-0,37	0,02	0,94	0,91	0,79
VALR3	0,04	0,02	0,95	0,93	0,79
VALR4	-0,09	0,02	1,04	1,04	0,76
VALR5	0,18	0,02	1	0,98	0,78
INTR1	0,29	0,02	1,02	1,02	0,76
INTR2	-0,04	0,02	0,93	0,91	0,77
INTR3	-0,15	0,02	0,92	0,9	0,77
INTR4	0,44	0,02	1,14	1,13	0,73
INTR5	-0,11	0,02	0,96	0,94	0,76
INTR6	-0,44	0,02	1,02	0,99	0,74
RELP1	0,22	0,02	1,03	1,01	0,77
RELP2	-0,5	0,02	0,95	0,91	0,77
RELP3	0	0,02	0,93	0,89	0,78
RELP4	0,38	0,02	1,12	1,14	0,76
RELP5	-0,1	0,02	0,94	0,89	0,78
TRDF1	-0,54	0,02	1,05	1,04	0,75
TRDF2	0,21	0,02	0,93	0,91	0,77
TRDF3	0,24	0,02	1,04	1,02	0,75
TRDF4	0,09	0,02	0,98	0,97	0,77
CLOR1	0,14	0,02	0,97	0,96	0,79
CLOR2	0,15	0,02	0,97	0,95	0,79
CLOR3	0,02	0,02	0,98	0,96	0,79
CLOR4	-0,16	0,02	0,98	0,96	0,78
CLOR5	-0,15	0,02	0,98	0,95	0,78
CLOR6	0,09	0,02	0,98	0,96	0,79
CLOR7	0,09	0,02	1,07	1,05	0,78
CLOR8	-0,17	0,02	1,01	0,98	0,78
ENTC1	-0,18	0,02	0,9	0,89	0,75
ENTC2	-0,1	0,02	0,91	0,88	0,75
ENTC3	0,29	0,02	1,06	1,05	0,73
ENTC4	0,26	0,02	1,14	1,12	0,72
ENTC5	-0,26	0,02	0,97	0,94	0,73
APPR1	-0,04	0,02	1	0,99	0,78
APPR2	0,15	0,02	1,01	1,01	0,77
APPR3	0,13	0,02	1,02	1,02	0,77
APPR4	0	0,02	0,94	0,92	0,79
APPR5	-0,24	0,02	1	0,98	0,77
TRPR1	0	0,03	0,95	0,91	0,84
TRPR2	0,09	0,03	0,93	0,89	0,85
TRPR3	0,03	0,02	1,13	1,08	0,81
TRPR4	-0,11	0,02	0,95	0,89	0,84
EVAL1	-0,27	0,02	0,86	0,85	0,75
EVAL2	0,34	0,02	0,87	0,87	0,76
EVAL3	-0,02	0,02	0,9	0,89	0,74
EVAL4	0,83	0,02	1,02	1,03	0,73
EVAL5	-0,42	0,02	0,96	0,95	0,71
EVAL6	-1,01	0,02	1,22	1,19	0,62
EVAL7	0,55	0,02	1,15	1,17	0,68

Tabla 3 Estadísticos descriptivos para la faceta ítems.

		ITEMS				
		Med	S.E.	InfitMS	OutfitMS	PtMea
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 4 Estadísticos descriptivos para la faceta Docentes y Estudiantes.

		DOCENTES					ESTUDIANTES				
		Med	S.E.	InfitMS	OutfitMS	PtMea	Med	S.E.	InfitMS	OutfitMS	PtMea
VALR E1	Media	2,29	0,16	1,11	1,11	0,71	-0,19	0,70	0,98	0,98	0,26
	SD	1,38	0,08	0,28	0,30	0,16	2,53	0,33	0,75	0,76	0,40
	Q1	1,44	0,12	0,96	0,93	0,69	-1,69	0,42	0,51	0,48	0,00
	Q3	2,77	0,19	1,29	1,31	0,78	1,47	0,85	1,30	1,33	0,57
INTR E2	Media	2,18	0,14	1,13	1,12	0,70	0,00	0,62	0,99	0,99	0,28
	SD	1,21	0,07	0,31	0,32	0,16	2,41	0,34	0,75	0,75	0,38
	Q1	1,45	0,10	0,96	0,95	0,67	-1,39	0,36	0,46	0,45	0,00
	Q3	2,84	0,17	1,36	1,35	0,76	1,49	0,72	1,40	1,42	0,60
RELP E3	Media	3,71	0,18	1,09	1,08	0,76	0,00	0,85	0,95	0,95	0,27
	SD	3,27	0,09	0,35	0,39	0,09	3,41	0,52	0,76	0,80	0,39
	Q1	2,31	0,11	0,86	0,83	0,72	-1,96	0,47	0,46	0,43	0,00
	Q3	4,23	0,22	1,29	1,33	0,82	1,72	0,97	1,14	1,16	0,63
TRDF E4	Media	1,19	0,16	1,20	1,20	0,72	0,00	0,69	0,98	0,97	0,26
	SD	0,96	0,08	0,31	0,32	0,08	2,29	0,31	0,78	0,78	0,45
	Q1	0,60	0,11	1,02	0,98	0,67	-1,35	0,44	0,43	0,42	0,00
	Q3	1,74	0,19	1,37	1,39	0,77	1,32	0,87	1,40	1,41	0,63
CLOR E5	Media	3,05	0,13	1,09	1,09	0,71	0,00	0,64	0,95	0,95	0,19
	SD	2,40	0,07	0,33	0,36	0,19	2,98	0,43	0,73	0,75	0,41
	Q1	1,81	0,09	0,87	0,83	0,67	-1,77	0,35	0,49	0,45	0,00
	Q3	3,82	0,16	1,28	1,32	0,79	-0,06	0,57	0,91	0,88	0,10
ENTC E6	Media	2,56	0,16	1,08	1,07	0,70	0,00	0,72	0,97	0,96	0,26
	SD	2,57	0,08	0,32	0,34	0,11	2,73	0,45	0,73	0,74	0,41
	Q1	1,44	0,10	0,88	0,86	0,65	-1,60	0,40	0,47	0,46	0,00
	Q3	3,18	0,19	1,27	1,28	0,76	1,38	0,81	1,27	1,28	0,59
APPR E7	Media	2,63	0,16	1,13	1,14	0,68	0,00	0,72	0,95	0,95	0,23
	SD	1,32	0,08	0,31	0,33	0,22	2,65	0,37	0,78	0,79	0,42
	Q1	1,77	0,11	0,90	0,90	0,65	-1,52	0,42	0,44	0,43	0,00
	Q3	3,67	0,19	1,32	1,32	0,78	1,57	0,83	1,26	1,28	0,56
TRPR E8	Media	2,67	0,20	1,19	1,15	0,78	0,00	0,90	0,91	0,91	0,12
	SD	1,51	0,10	0,42	0,42	0,10	3,42	0,40	0,98	0,99	0,47
	Q1	1,89	0,14	0,84	0,80	0,73	-2,22	0,55	0,03	0,03	0,00
	Q3	3,23	0,25	1,46	1,39	0,85	2,25	1,25	1,00	1,02	0,44
EVAL E9	Media	0,77	0,11	1,06	1,06	0,68	0,00	0,49	0,97	0,97	0,31
	SD	0,72	0,05	0,25	0,25	0,10	1,88	0,29	0,69	0,69	0,35
	Q1	0,45	0,08	0,88	0,88	0,64	-0,92	0,30	0,43	0,44	0,00
	Q3	1,09	0,13	1,25	1,25	0,73	0,09	0,45	0,86	0,86	0,34

Tabla 5. Confiabilidad de Separación (Rasch) para cada faceta y dimensión considerada.

	Items		Docentes		Estudiantes		
	Confiabilidad Separación	Prueba Efectos Fijos (todos iguales) valor p	Confiabilidad Separación (Rasch)	Prueba Efectos Fijos (todos iguales) valor p	Confiabilidad Separación (Rasch)	Confiabilidad entre Evaluadores (Rasch)	Prueba Efectos Fijos (todos iguales) valor p
VALR (E1)	0,99	<0,005	0,98	<0,005	0,91	0,09	<0,005
INTR (E2)	0,99	<0,005	0,98	<0,005	0,91	0,09	<0,005
RELP (E3)	0,99	<0,005	1,00	<0,005	0,92	0,08	<0,005
TRDF (E4)	0,99	<0,005	1,00	<0,005	0,92	0,08	<0,005
CLOR (E5)	0,97	<0,005	1,00	<0,005	0,93	0,07	<0,005
ENTC (E6)	0,99	<0,005	1,00	<0,005	0,90	0,10	<0,005
APPR (E7)	0,98	<0,005	0,98	<0,005	0,91	0,09	<0,005
TRPR (E8)	0,88	<0,005	0,98	<0,005	0,92	0,08	<0,005
EVAL (E9)	1	<0,005	0,97	<0,005	0,91	0,09	<0,005

5.3. Análisis de 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 5.2. En esta sección 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. Análisis de Interacciones (sesgos) entre facetas para cada dimensión considerada.

	Curso*Estudiante		Curso*Item		Docente*Item		Docene*Estudiante		Estudiante*Item	
	Var. Exp. %	Prueba de Efectos fijos (todos cero) Valor p	Var. Exp. %	Prueba de Efectos fijos (todos cero) Valor p	Var. Exp. %	Prueba de Efectos fijos (todos cero) Valor p	Var. Exp. %	Prueba de Efectos fijos (todos cero) Valor p	Var. Exp. %	Prueba de Efectos fijos (todos cero) Valor p
VALR(E1)	9,86	<0,01	0,57	<0,01	0,86	<0,01	18,52	<0,01	11,29	1
INTR(E2)	10,37	<0,01	0,76	<0,01	1,22	<0,01	19,13	<0,01	12,29	1
RELP(E3)	11,46	<0,01	0,32	<0,01	0,65	<0,01	20,19	<0,01	7,69	1
TRDF(E4)	10,63	<0,01	1,16	<0,01	1,61	<0,01	17,27	<0,01	13,42	1
CLOR(E5)	9,78	<0,01	0,14	<0,01	0,66	<0,01	17,77	<0,01	9,37	1
ENTC(E6)	10,50	<0,01	0,35	<0,01	1,27	<0,01	19,89	<0,01	10,84	1
APPR(E7)	12,63	<0,01	0,39	<0,01	0,67	<0,01	21,86	<0,01	9,82	1
TRPR(E8)	11,84	<0,01	0,12	<0,01	0,39	1	11,50	<0,01	16,62	0,01
EVAL(E9)	7,19	<0,01	1,12	<0,01	1,53	<0,01	7,07	0,74	30,36	1

En la Tabla 6 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, considerando las interacciones estadísticamente significativas, el porcentaje de la varianza explicada es mayor en general para la interacción Docente*Estudiante, seguida de la interacción Curso*Estudiante. Las demás interacciones, en general no son estadísticamente significativas, o bien, el porcentaje de la varianza explicada por la interacción es pequeño.

Tabla 7. Análisis de la invarianza de las medidas. $r(1,2)$ representa la correlación lineal entre dos sub-muestras aleatorias complementarias, mientras que $r(1,C)$ y $r(2,C)$ representan las correlaciones entre cada sub-muestra y la muestra completa. Los valores de correlación lineal se reportan como r (r corregido para atenuación).

Sub-escala	Correlación entre muestras	Invarianza respecto a Faceta Estudiantes		Invarianza respecto a Faceta 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)

Más allá del patrón general encontrado, que resulta consistente con las predicciones teóricas realizadas, se observa 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 explica un 16,62% de la varianza y es 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 resulta no estadísticamente significativa y explica 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.

5.4. Análisis de la 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 7 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. 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 7).

6. CONCLUSIONES Y CONTRIBUCIONES

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 de diagnósticos y de mejora de la calidad de la enseñanza.

BIBLIOGRAFÍA

- 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.
- Braskamp, L. A., y Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. Jossey-Bass higher and adult education series. ERIC.
- Centra, J. A. (1976). The influence of different directions on student ratings of instruction. *Journal of Educational Measurement*, 13 (4), 277-282.
- Dwinell, P. L., y Higbee, J. L. (1993). Students' perceptions of the value of teaching evaluations. *Perceptual and Motor Skills*, 76 (3), 995-1000.
- 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.
- Marlin Jr, J. W. (1987). Student perceptions of end-of-course evaluations. *The Journal of Higher Education*, 704-716.
- 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.
- Seldin, P. (1989). How colleges evaluate professors. 1988 vs. 1983. *AAHE Bulletin*, 41 (7), 3-7.
- Theall, M., Abrami, P. C., y Mets, L. A. (2001). The student ratings debate: Are they valid? how can we best use them?