

DISEÑO DE UN MICRO-INSTRUMENTO PARA MEDIR EL CLIMA DE APRENDIZAJE EN CUESTIONARIOS DE CONTEXTO¹

Design of a micro-instrument for the measurement of learning climate useful to develop context questionnaires

EMELINA LÓPEZ-GONZÁLEZ
Universidad de Valencia

JAVIER TOURÓN
Universidad de Navarra

FRANCISCO JAVIER TEJEDOR
Universidad de Salamanca

En el presente artículo se trabaja a partir de una escala base como cuestionario de contexto que mide el constructo clima de aprendizaje. El interés se centra en simplificar la medida de dicho constructo en las evaluaciones de sistemas educativos. El estudio se lleva a cabo con dos muestras de estudiantes de sexto de educación primaria y cuarto de educación secundaria obligatoria, respectivamente. El objetivo es diseñar un micro-instrumento parsimonioso, fiable, válido y con suficiente capacidad para discriminar a los alumnos. Para ello se determina una versión reducida de dicha escala base utilizando el procedimiento de reducción R-AVACO. El diseño se lleva a cabo en dos fases. En una primera se seleccionan los ítems de mayor calidad y de mayor poder discriminativo, empleando para ello diversas técnicas estadísticas: análisis exploratorio y estudio de observaciones ausentes, análisis de conglomerados y análisis de fiabilidad. En una segunda fase se estiman las puntuaciones totales, se explora la versión reducida de la escala base y se realiza la validación del micro-instrumento. Como resultado, se obtienen dos versiones reducidas de la escala base inicial, adaptadas a las muestras con las que se trabaja. Estos micro-instrumentos terminan siendo valorados siguiendo los criterios de eficiencia, funcionalidad y eficacia.

Palabras clave: *Evaluación de sistemas educativos, Clima de aprendizaje, Cuestionario de contexto, Fiabilidad, Validez.*

Introducción

La evaluación de un sistema educativo supone un momento culminante de intervención de un Gobierno en el ámbito educativo con objeto de

revisar la respuesta que dicho sistema debe dar a las necesidades sociales. Su finalidad última es la toma de decisiones sobre preguntas tan esenciales como: qué, cómo, a quién y para qué educar. Si tal evaluación es realmente trascendental, su

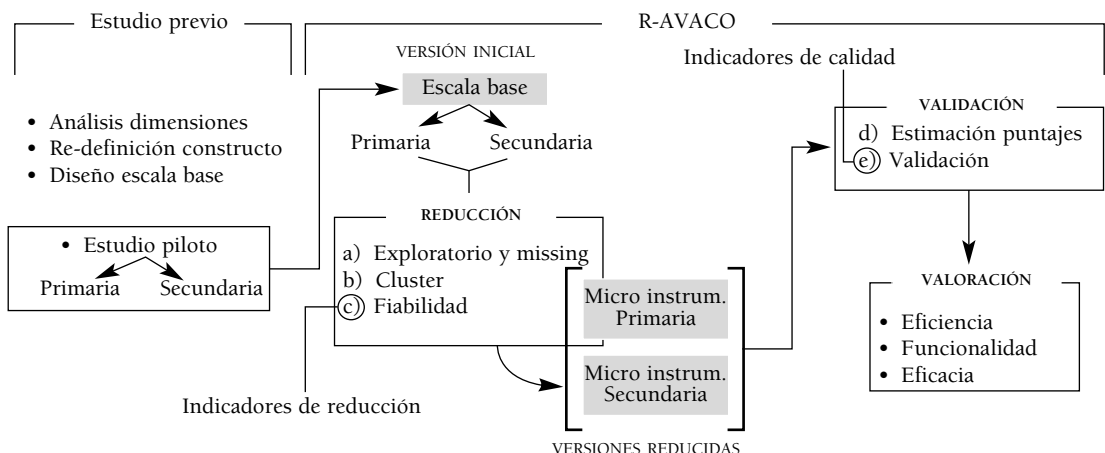
ejecución es similar en su complejidad, y ello debido al gran número de aspectos, factores, dimensiones y variables que considerar. Por lo general tales variables se agrupan en cuatro categorías, a saber: variables de entrada, de proceso, de contexto y de producto. Los planes de evaluación de sistemas educativos han atendido especialmente a variables de producto (resultado escolar), las cuales, aun siendo muy necesarias, resultan insuficientes para conocer en profundidad la situación del sistema, especialmente para identificar los elementos que hay que mejorar —algo imprescindible—, así como para potenciar la utilidad de la evaluación misma.

Para este cometido, un aspecto que requiere especial atención es el diseño de instrumentos orientados a la recogida de información de variables de entrada, proceso y contexto, los comúnmente denominados *cuestionarios de contexto*. Así, por ejemplo, las evaluaciones relativas al proceso (estilo de enseñanza, metodología didáctica, clima de aprendizaje, clima social en el aula, etc.), por su complejidad y dificultad de abordaje, no terminan de estar bien resueltas, aun cuando son variables de las que se conoce su funcionalidad y para las que existen soluciones metodológicas que posibilitan su evaluación. Por esta razón, toda la atención que se preste al

proceso de construcción de cuestionarios de contexto es, sin duda, garantía de una mejora trascendental para la evaluación del sistema educativo. Entre otras cuestiones, cabe atender a la revisión sustantiva de los constructos a los que se refieren los cuestionarios de contexto; el diseño de instrumentos y escalas acordes con las dimensiones internas del constructo; la recogida de las variables contenidas en cada dimensión atendiendo a sus propiedades métricas, y la identificación de los ítems, elementos o reactivos clave que faciliten la fiabilidad y validez necesarias.

Tal ha sido precisamente el interés de un proyecto de investigación que trabaja constructos de las cuatro categorías señaladas arriba¹. El estudio que aquí presentamos se articula dentro de ese proyecto y tiene como objeto simplificar la medida de una de las variables de contexto —el constructo clima de aprendizaje— a partir de un instrumento elaborado en tal proyecto. Es decir, en un momento anterior se estudiaron las dimensiones y reactivos utilizados en otras escalas relacionadas con el clima de aprendizaje, se redefinió el constructo en colaboración con un comité de expertos, se diseñó una escala de clima de aprendizaje y se empleó dicha escala en un estudio piloto². Un esquema del proceso se representa en la figura 1.

FIGURA 1. Esquema del proceso de reducción R-AVACO



Nuestro trabajo comienza con la escala (a partir de ahora *escala base*) que resulta del estudio previo. En ella se ha aplicado la solución metodológica de reducción expuesta en Jornet, González-Such y Perales (2012) —denominada R-AVACO—, diseñada de forma explícita para tratar cuestionarios de contexto de los diversos constructos o variables que acontecen en la evaluación de un sistema educativo (ejemplos de estos instrumentos se recogen en Pérez-Carbonell, Ramos y López-González, 2009 y Biencinto, González-Barbera, García-García, Sánchez-Delgado y Madrid, 2009). Es decir, se construye un micro-instrumento a partir de la reducción de una escala base que evalúa el clima de aprendizaje, con el interés de identificar los reactivos o ítems clave, a saber, los mejores predictores de las puntuaciones totales de las dimensiones internas de la escala base (figura 1).

Siguiendo el principio de parsimonia, uno de los objetivos es conseguir instrumentos que recojan la máxima información con el menor número de ítems y hacerlo, además, de forma fiable y válida. Esta menor longitud de los cuestionarios reduce el tiempo de su cumplimentación y permite disminuir también el número de no respuestas. No perdamos de vista que la evaluación de sistemas educativos precisa de numerosos instrumentos aplicados en un reducido número de sesiones, instrumentos que deben posibilitar la recogida de las diversas variables contenidas en las dimensiones de contexto y de producto. Por este motivo, cualquier solución metodológica que pretenda configurar un cuestionario de contexto ha de optimizar la calidad del instrumento, identificando los ítems con mayor poder de discriminación. Pero no solo la menor extensión de las escalas es una característica deseable. Los ítems deben ser realmente representativos del constructo objeto de medición. En suma, se trata de conseguir micro-instrumentos con las siguientes características: 1) fiabilidad, 2) poder de clasificación de los sujetos, 3) validez de constructo y 4) eficiencia (parsimoniosos y de rápida administración). Por consiguiente, el objetivo último de este trabajo es, además de efectuar la reducción adecuada de

la escala base del constructo clima de aprendizaje, hacerlo atendiendo a estos *criterios de diseño*.

Método

La reducción R-AVACO se concreta en los *pasos* que de manera esquemática se muestran en la figura 1: a) análisis exploratorio y estudio de valores ausentes (*missing values*) según la prueba MCAR³ de Little y Rubin (2002), b) análisis de conglomerados, c) análisis de fiabilidad, d) estimación de la puntuación total y exploración de la versión reducida y e) validación final del micro-instrumento. Cada uno de estos análisis aporta elementos importantes para juzgar la oportunidad de eliminar paulatinamente los ítems o reactivos que no cumplan suficientemente con el criterio estadístico fijado en cada momento. Estos análisis son aplicados sobre los datos obtenidos a partir de la escala base, cuyos ítems se recogen en la tabla 1.

Participantes

Se trabajó con alumnos distribuidos en dos muestras: 310 alumnos de sexto de primaria (11-12 años) y 179 alumnos de cuarto de secundaria (15-16 años). Fueron tomados los datos de siete centros: tres concertados y dos públicos para la muestra de primaria y dos centros concertados para la muestra de secundaria. Los alumnos no fueron seleccionados de forma probabilística, sino accidental.

Instrumentos

Se aplicó el mismo cuestionario base a todos los casos. Por medio de la reducción R-AVACO se diseñaron dos micro-instrumentos o versiones reducidas de la escala base, uno para cada nivel de estudios por tratarse de muestras con características diferenciadas. No se ha pretendido una comparación de resultados, sino obtener un instrumento final adaptado a cada nivel que cumpla con los criterios de diseño referidos anteriormente.

Los análisis estadísticos se han efectuado con los programas SPSS™ 19 y Excel™ 2007.

Resultados

La tabla 1 contiene el significado de los ítems empleados en la escala base. Se ha respetado la numeración original con la que fueron presentados para ser respondidos. Los elementos aparecen agrupados en dimensiones sustantivas. Para la muestra de primaria se informa del número de respuestas, el número de observaciones ausentes, el porcentaje que estas últimas representan y el conglomerado o *cluster* en el que queda situado cada ítem en el análisis de conglomerados que veremos más tarde.

a) Estudio de valores ausentes (*missing values*) y exploración de ítems:

Al llevar a cabo el recuento de valores ausentes en la muestra de primaria, se observa que no hay elementos con un 5% o más de no respuestas. Aquellos ítems en los que se ha encontrado una mayor presencia no superan el 3,2% (tabla 1). No obstante, si imputamos las observaciones ausentes por la media estimada de la serie correspondiente, se producen diferencias significativas con la prueba *t* de Student entre las parejas de ítems (presente/ausente). Igualmente, la prueba MCAR aplicada a las matrices de covarianzas y correlaciones aporta estadísticos J_i cuadrado significativos ($\chi^2 = 536.08$; $gl = 406$ y $p < .001$), lo que pone en duda el carácter aleatorio de las no respuestas y sugiere la conveniencia de su imputación. En esta misma línea, al hacer el análisis de las observaciones ausentes por casos se aprecia que los sujetos con no respuestas ocasionan valores perdidos en numerosos elementos (ítems 17, 18 y 28 a 34), lo que induce a pensar que quizá sea más adecuado considerar solo sujetos con información completa. Si es así, trabajaríamos en realidad con una muestra de primaria de 300 casos en lugar de los 310 iniciales, perdiendo un 3,22% de alumnos.

Las consecuencias de la eliminación de estos sujetos en los análisis posteriores no son importantes. Tal es el caso del análisis de conglomerados (*cluster analysis*) que efectuamos después. Es sabido que los datos ausentes crean más o menos problemas dependiendo del tipo de análisis que se emplee. Con el análisis factorial y el análisis de fiabilidad, por ejemplo, su presencia afecta porque de entrada se emplean matrices de correlaciones, lo que no sucede con el análisis de conglomerados y el escalamiento multidimensional, al utilizar estas distancias euclídeas y no correlaciones. Este argumento, junto con el dato de que los valores ausentes no superan el 5% en ningún ítem, nos anima finalmente a seguir trabajando con la muestra completa de primaria (310 casos), imputando la mediana para sustituir los valores ausentes (nos parece conveniente este estadístico y no la media por tratarse de respuestas en escala Likert).

En la muestra de secundaria no hay elementos con un 1% o más de observaciones ausentes. Al imputar las observaciones ausentes por la media estimada, no aparecen diferencias significativas con la prueba *t* de Student entre las parejas de ítems (presente/ausente). La prueba MCAR aplicada a las matrices de covarianzas y correlaciones no aporta estadísticos J_i cuadrado significativos ($\chi^2 = 123.6$; $gl = 116$ y $p = .36$), lo que confirma el carácter aleatorio de los valores ausentes. En consecuencia, en secundaria no eliminamos ningún ítem según este criterio y pasamos a sustituir las no respuestas por la mediana del elemento correspondiente.

Con el análisis exploratorio se hace efectiva una segunda reducción. El comportamiento general de los ítems se explora por medio de diagramas de caja. El objetivo es prescindir de aquellos elementos que tienen un comportamiento anómalo (especialmente falta de varianza). En ambas muestras hay algunos elementos que carecen de la variabilidad suficiente: los ítems 17, 18 y 23 en la muestra de primaria y los elementos 14, 15, 16 y 21 en la de secundaria.

TABLA 1. Elementos de la escala base. Valores ausentes en la muestra de primaria

Ítem significado	N	Primaria		Anál. cluster >1% Conglom.
		Presencia de missing	Porcentaje missing	
Dimensión 1: interés y preocupación				
14 Los profesores se interesan por lo que aprendemos	310	0	0.0	1
15 Me parece interesante lo que aprendemos en las clases	310	0	0.0	1
16 Pongo interés en los trabajos y actividades de clase	308	2	0.6	1
17 La mayor parte de mis compañeros quieren aprender	306	4	1.3	X 3
18 A algunos profesores les da lo mismo que trabajemos bien o mal	306	4	1.3	X 2
19 Algunas clases me interesan muy poco porque creo que no son importantes y me hacen perder tiempo	308	2	0.6	2
Dimensión 2: satisfacción				
20 Me siento orgulloso de mi trabajo en clase	309	1	0.3	1
21 Estoy satisfecho con lo que aprendo en clase	308	2	0.6	1
22 En algunas clases me aburro porque tratan cosas ya conocidas o demasiado fáciles	308	2	0.6	2
23 Los profesores están contentos cuando ven que aprendemos y trabajamos bien	309	1	0.3	1
Dimensión 3: expectativas				
24 Los profesores quieren que aprendamos más cosas de las que podemos aprender	309	1	0.3	3
25 La mayoría de los compañeros estamos satisfechos con lo que aprendemos en clase	307	3	1.0	3
26 Al final de curso voy a tener un buen nivel para continuar en los cursos siguientes	307	3	1.0	1
27 Lo que estoy aprendiendo me va a servir para el futuro	308	2	0.6	1
Dimensión 4: competitividad				
28 En clase, siempre hay compañeros que quieren quedar mejor en las tareas que los demás	300	10	3.2	X 3
29 La mayor parte de los compañeros quieren trabajar tan bien como lo hacen los mejores de la clase	300	10	3.2	X 3
Dimensión 5: cooperación				
30 Normalmente los compañeros nos ayudamos para hacer las tareas de clase	301	9	2.9	X 3
31 La mayor parte de mis compañeros dejan trabajar a los demás	300	10	3.2	X 3
Dimensión 6: ambiente físico				
32 Hay ruido en el aula	301	9	2.9	X 2
33 Hay desorden en el aula	301	9	2.9	X 2
34 La distribución de la clase permite trabajar a gusto	302	8	2.9	X 3

b) Análisis de conglomerados:

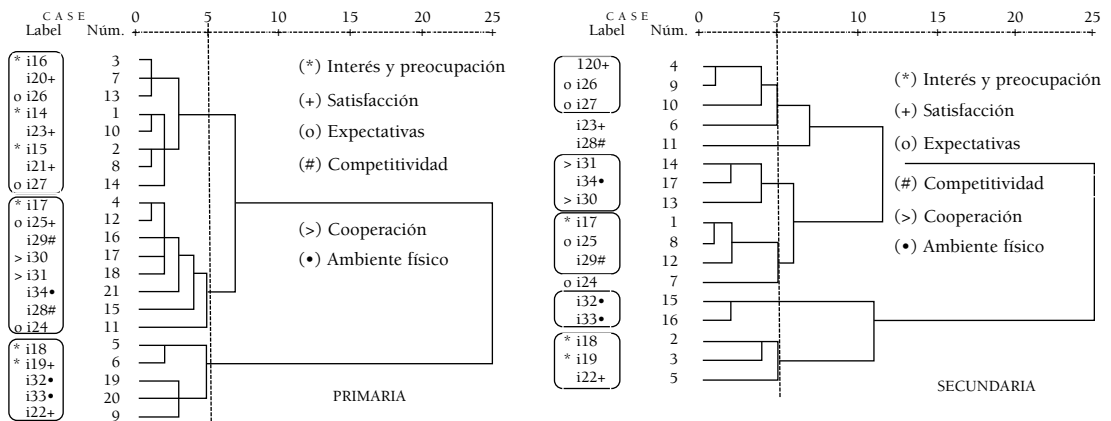
Los dos análisis de conglomerados que se llevan a cabo tienen como objetivo explorar las dimensiones sustantivas que se emplearon para elaborar la escala base. Se efectúa una opción jerárquica con el método Ward de conglomeración, usando como medida de disimilaridad la distancia euclídea por tratarse de escalas ordinales (tal y como sugieren López González, Pérez Carbonell y Ramos, 2011). En la figura 2 se muestran los dendrogramas resultantes. Siguiendo la recomendación de R-AVACO, se consideran las agrupaciones construidas según una distancia combinada o re-escalada igual o menor a cinco (Jornet *et al.*, 2012). Las primeras agrupaciones se establecen con una distancia más pequeña, lo que indica mayor similitud entre los elementos.

En la muestra de primaria puede hablarse de tres conglomerados. Como se observa en la figura 2 y en las tablas 1 y 4, en el primer conglomerado se sitúan especialmente elementos de las dimensiones *interés y preocupación* (14, 15 y 16) y *satisfacción* (20, 21, 23); en la segunda agrupación destacan las dimensiones *cooperación* (30 y 31) y *competitividad* (28 y 29); en

el tercer agrupamiento se asocian elementos de las dimensiones *interés y preocupación* (18 y 19) y *ambiente físico* (32 y 33). Según estas configuraciones podría hablarse de una cierta correspondencia entre las dimensiones sustantivas y las que se configuran a partir de los datos. En cualquier caso, veremos más adelante cómo en esta muestra de primaria serán los elementos agrupados en el primer conglomerado los que superan los indicadores de reducción para formar parte de la versión reducida.

En la muestra de secundaria, aunque se perfilan cuatro agrupaciones de varios ítems con menor distancia re-escalada, en realidad la correspondencia entre la configuración sustantiva de dimensiones teóricas y la observada se establece entre las siguientes parejas de ítems: 26 y 27 de *expectativas*, 18 y 19 de *interés y preocupación*, 30 y 31 de *cooperación* y 32 y 33 de *ambiente físico* (ver figura 2 y tabla 5). Estas tres últimas parejas se encuentran también agrupadas en la muestra de primaria, por lo que puede afirmarse que son los reactivos que, según el punto de vista sustantivo, mejor han sido percibidos por los alumnos como asociados con las dimensiones teóricas dentro de las cuales fueron ubicados.

FIGURA 2. Dendrogramas de los análisis cluster



c) Análisis de fiabilidad:

El doble objetivo de este tercer paso en la R-AVACO es la selección de los elementos que mejor se relacionan con la puntuación total de la escala, así como optimizar dicha información eliminando los ítems que resultan redundantes. Ambos procesos se llevan a cabo por medio de sucesivos análisis de fiabilidad con el modelo alfa de Cronbach.

De forma secuencial se van tomando decisiones a partir de la información de tres *indicadores de reducción*: 1) suprimir los elementos cuyo indicador *alfa si se elimina el elemento* informa de un incremento de la fiabilidad, 2) prescindir de los reactivos con correlación ítem-total menor a .30 ($r_{it-i} < .30$) y 3) eliminar los ítems que, aunque tengan asociada una correlación con otro

elemento superior a .50 en la matriz de correlaciones entre ítems ($r_{xy} > .50$), posean al mismo tiempo una peor correlación r_{it-i} .

Ahora bien, este último indicador se utiliza considerando la función sustantiva de los ítems, de tal manera que se mantienen aquellos elementos que pertenecen a distintas dimensiones del constructo (tabla 1), aunque debieran ser eliminados según el indicador de reducción correspondiente. Como veremos, esto sucede en la muestra de primaria con los ítems 30, 31 y 34 y en secundaria con el reactivo 17.

En la tabla 2 se reflejan los resultados de los tres análisis de fiabilidad una vez aplicados los indicadores de reducción. En la muestra de primaria quedan seleccionados diez elementos

TABLA 2. Análisis de fiabilidad

	Primaria		Secundaria			
Primer análisis	$\alpha = .374$		$\alpha = .374$			
	Se eliminan ítems: 19, 22, 24, 28, 32, 33		Se eliminan: 18, 19, 22, 32, 33			
Segundo análisis	$\alpha = .733$		$\alpha = .672$			
	Se eliminan ítems: 27, 29		Se eliminan: 17, 24, 28, 30			
Tercer análisis	$\alpha = .734$		$\alpha = .678$			
	Nº de elementos: 10		Nº de elementos: 9			
	ítem	(1)	(2)	ítem	(1)	(2)
	i14	.413	.711			
	i15	.491	.699			
	i16	.388	.714			
				i17	.233	.674
	i20	.423	.709	i20	.307	.661
	i21	.416	.709			
				i23	.428	.637
	i25	.464	.702	i25	.494	.626
	i26	.276	.729	i26	.434	.633
				i27	.350	.654
				i29	.301	.662
	i30	.362	.719			
	i31	.386	.715	i31	.358	.650
	i34	.372	.719	i34	.304	.665

(1) Correlación ítem-total corregida. (2) Alfa de Cronbach si se elimina el elemento.

y en la de secundaria nueve. Los reactivos comunes a las dos muestras son: 20, 25, 26, 31 y 34. Los índices alfa de Cronbach finales se calculan con los ítems no eliminados y resultan ser bastante elevados (.73 y .67), lo que permite confiar en la selección efectuada.

Las tablas 4 y 5 resumen el comportamiento de los ítems de las distintas dimensiones sustantivas una vez aplicados los tres primeros pasos de la reducción R-AVACO. El conjunto de reactivos seleccionados constituye ya la versión reducida de la escala base sobre la que aplicamos los análisis de las fases posteriores (en las tablas 6 y 7 se recoge el contenido de los ítems). Puede apreciarse que la reducción es importante, especialmente a partir de los indicadores utilizados en este tercer paso de la R-AVACO. En la muestra de primaria, por ejemplo, se eliminan ocho elementos (19, 22, 24, 27, 28, 29, 32, 33). En la de secundaria se suprimen nueve (17, 18, 19, 22, 24, 28, 29, 32, 33). El recuento global queda recogido en la tabla 3.

Veamos con más detalle cómo se ven afectadas las dimensiones a partir del comportamiento de los ítems. En la muestra de primaria (tabla 4) desaparece de la versión reducida la dimensión *competitividad* al eliminar los elementos 28 y 29. Desaparecen también todos los ítems del tercer conglomerado del análisis *cluster* anterior (recordemos la figura 2). Sin embargo, excepto para el ítem 27, quedan seleccionados todos los reactivos que se agruparon en el primer

conglomerado. Por otro lado, decidimos mantener en la versión reducida los ítems 30, 31 y 34 ya que superan los indicadores de reducción, aunque cumplan el criterio de eliminación por presencia de valores ausentes, lo cual no debe atribuirse a la calidad del elemento sino al comportamiento de los sujetos en sus respuestas. De este modo se conservan las dimensiones *cooperación* y *ambiente físico* para formar parte de la versión reducida.

En la muestra de secundaria todos los elementos de la dimensión *interés y preocupación* que funcionaron bien en la exploración de dimensiones del análisis de conglomerados quedan ahora eliminados. Se fuerza a mantener el ítem 17 con objeto de que esta dimensión esté presente en la versión reducida. Igualmente, de las cuatro parejas de ítems cuya agrupación coincidía con las dimensiones teóricas según el análisis de conglomerados (tabla 5), solo superan los indicadores de reducción la pareja de ítems 26 y 27, de *expectativas*, y el ítem 31. La selección de este último elemento permite mantener la dimensión *cooperación*. En conjunto, todas las dimensiones sustantivas de la escala base han podido conservarse para formar parte del micro-instrumento.

d) Estimación de la puntuación total y exploración de la versión reducida:

La puntuación de la versión reducida se estima como la suma total de valores para cada elemento.

TABLA 3. Recuento de eliminación de elementos

	Muestra	% valores ausentes	Exploratorio	Fiabilidad	Total eliminados	Total seleccionados
Versión inicial: escala base de 21 elementos	Primaria	9	3	8	10 de 21	11
		Versión reducida o micro-instrumento de 11 elementos				
	Secundaria	0	4	9	12 de 21	9
		Versión reducida o micro-instrumento de 9 elementos				

TABLA 4. Eliminación progresiva de ítems en primaria según los distintos pasos de la reducción R-AVACO

Dimensión sustantiva	Agrupados por cluster		% valores ausentes > 1%	Eliminados en exploratorio	Eliminados en fiabilidad	Seleccionados para versión reducida
	ítem	nº cluster				
Interés y preocupación	14	1				X
	15	1				X
	16	1				X
	17	3	X	No variabilidad		
	18	2	X	No variabilidad		
	19	2			Aumento de α	
Satisfacción	20	1				X
	21	1				X
	22	2			Aumento de α	
	23	1		No variabilidad		X
Expectativas	24	3			Aumento de α	
	25	3				X
	26	1				X
	27	1			$r_{it-i} < .30$	
Competitividad	28	3	X		Aumento de α	
	29	3	X		$r_{it-i} < .30$	
Cooperación	30	3	X			Dudoso
	31	3	X			Dudoso
Ambiente físico	32	2	X		Aumento de α	
	33	2	X		Aumento de α	
	34	3	X			Dudoso

Aunque este valor no ha sido empleado para la selección de ítems, se utiliza a partir de ahora como referente del nuevo instrumento, necesario para el proceso de validación del siguiente paso e).

La exploración de la calidad de la versión reducida permite que esta sea ajustada al micro-instrumento definitivo. Para ello se realizan varias exploraciones apoyadas en las ordenaciones que se obtienen de los análisis que se comentan a continuación y que pasamos a considerar como *indicadores de calidad*:

d. 1.) Orden o rango de los elementos atendiendo a su potencia discriminativa según los

resultados obtenidos en la prueba *t* de Student que contrasta el 27% inferior y superior⁴. Estos análisis resultan significativos muy probablemente como consecuencia del proceso de reducción de los elementos menos discriminativos realizado hasta ahora. Si no hubiera sido así, podrían emplearse los resultados de esta prueba para seleccionar ítems también en este momento.

d. 2.) Resultados según el análisis de fiabilidad que se efectúa con la versión reducida. Se combinan los siguientes indicadores estadísticos:

- Nivel de intensidad de la correlación del ítem con el test (r_{it-t}).

TABLA 5. Eliminación progresiva de ítems en secundaria según los distintos pasos de la reducción R-AVACO

Dimensión sustantiva	Agrupados por cluster	% valores ausentes > 1%	Eliminados en exploratorio	Eliminados en fiabilidad	Seleccionados para versión reducida
Interés y preocupación	14	No variabilidad			
	15	No variabilidad			
	16	No variabilidad			
	17			$r_{it-i} < .30$	X (¿?)
	18	Sí		Aumento de α	
	19	Sí		Aumento de α	
Satisfacción	20				X
	21	No variabilidad			
	22			Aumento de α	
	23				X
Expectativas	24			$r_{it-i} < .30$	
	25				X
	26	Sí			X
	27	Sí			X
Competitividad	28			$r_{it-i} < .30$	
	29			$r_{it-i} < .30$	X
Cooperación	30	Sí			
	31	Sí			X
Ambiente físico	32	Sí		Aumento de α	
	33	Sí		Aumento de α	
	34				X

- Fiabilidad del micro-instrumento eliminando el ítem. En este caso estimamos el rango de los elementos, entendiendo que el mejor valor de *alfa eliminado el ítem* resulta negativo respecto a la calidad del mismo.
- Valor de la suma de covarianzas de cada ítem con todos los demás elementos.

En cada uno de estos procesos se obtiene una ordenación que se combina finalmente calculando el rango medio de las tres ordenaciones. En las tablas 6 y 7 presentamos el orden definitivo de ítems. Este último rango constituye un elemento de referencia imprescindible para la validación posterior, en la que aquellos elementos con menor rango son los que pueden ser juzgados como de mayor calidad.

La lectura de las tablas 6 y 7 muestra una clara diferencia en las ordenaciones de elementos de las dos muestras. Si sabíamos que la composición de ambos micro-instrumentos solo coincidía en cinco elementos (ítems 20, 25, 26, 31 y 34), la ordenación de estos resulta también dispar. Por el contrario, sí parece existir mayor coincidencia en las dimensiones: en la muestra de primaria son algunos ítems de *interés y preocupación*, *expectativas* y *satisfacción* los que están mejor situados. En la de secundaria, sobre todo elementos de *expectativas* y *satisfacción*. Se colocan con un rango mayor los reactivos de las dimensiones *competitividad*, *cooperación* y *ambiente físico*. Si quisiéramos destacar algún elemento común para ambas muestras, sería el ítem 25

TABLA 6. Ordenación de los ítems del micro-instrumento de primaria según su calidad

Rango	Ítem	Contenido del elemento	Dimensión
1	15	Me parece interesante lo que aprendemos en las clases	Interés y preocupación
2	25	La mayoría de los compañeros estamos satisfechos con lo que aprendemos en clase	Expectativas
3	21	Estoy satisfecho con lo que aprendo en clase	Satisfacción
4	20	Me siento orgulloso de mi trabajo en clase	Satisfacción
5	34	La distribución de la clase permite trabajar a gusto	Ambiente físico
6	30	Normalmente los compañeros nos ayudamos para hacer las tareas de la clase	Cooperación
7	16	Pongo interés en los trabajos y actividades de clase	Interés y preocupación
8	31	Los profesores se interesan por lo que aprendemos	Cooperación
9	14	La mayor parte de mis compañeros quieren aprender	Interés y preocupación
10	26	Al finalizar el curso voy a tener un buen nivel para continuar en los cursos siguientes	Expectativas

(la mayoría de los compañeros estamos satisfechos con lo que aprendemos en clase), que en secundaria se coloca en primer lugar y en primaria en segundo.

e) Validación final del micro-instrumento: La validación del micro-instrumento se realiza por medio de un proceso iterativo que persigue estudiar su calidad global a tres niveles: 1) uso

TABLA 7. Ordenación de los ítems del micro-instrumento de secundaria según su calidad

Rango	Ítem	Contenido del elemento	Dimensión
1	25	La mayoría de los compañeros estamos satisfechos con lo que aprendemos en clase	Expectativas
2	23	Los profesores están contentos cuando ven que aprendemos y trabajamos bien	Satisfacción
3	26	Al finalizar el curso voy a tener un buen nivel para continuar en los cursos siguientes	Expectativas
4	27	Lo que estoy aprendiendo me va a servir si decido buscar trabajo al terminar la ESO	Expectativas
5	31	La mayor parte de mis compañeros dejan trabajar a los demás	Cooperación
6	34	La distribución de la clase permite trabajar a gusto	Ambiente físico
7	20	Me siento orgulloso de mi trabajo en clase	Satisfacción
8	29	La mayor parte de mis compañeros quieren trabajar tan bien como lo hacen los mejores de la clase	Competitividad
9	17	La mayor parte de mis compañeros quieren aprender	Interés y preocupación

de la puntuación total, 2) capacidad del instrumento para clasificar centros, aulas o alumnos y 3) mantenimiento de las características sustantivas de la escala base inicial. Para ello interesa comprobar la equivalencia de las puntuaciones globales de las dos versiones del instrumento —la escala base y el micro-instrumento—, y la equivalencia también entre las clasificaciones de alumnos, aulas y/o centros que se obtienen en ambas versiones. Con los ítems de las versiones reducidas (tablas 5 y 6) se procede a los siguientes análisis:

e. 1.) Correlacionar los totales de ambas versiones. Siguiendo la recomendación de R-AVACO, las iteraciones se paran en este punto si la correlación entre las versiones es inferior a .90. En primaria se obtiene una $r = .79$ significativa y en secundaria $r = .736$, significativa también. En ambas muestras no se cumple el criterio fijado, lo que aconseja no continuar con la reducción.

e. 2.) Comparar las clasificaciones de la escala base y del micro-instrumento por medio de J_i cuadrado: en ambas muestras las clasificaciones resultan equivalentes.

e. 3.) Comparados por centros (o por distintas unidades muestrales —centros y aulas—). Como criterio para representar el micro-instrumento se toma el porcentaje de unidades en las que se produce una clasificación concurrente entre ambas versiones. Así, en las dos muestras se cumple el criterio de clasificaciones equivalentes por centros, aunque se producen casos con frecuencias en las casillas menores que cinco. Según la R-AVACO, el proceso iterativo debe detenerse cuando aparezcan asociaciones de Spearman $r < .70$, lo que sucede en dos colegios de primaria y también en dos centros de secundaria.

e. 4.) Revisar el criterio sustantivo. La estructura inicial sustantiva de dimensiones en la muestra de primaria quedó mermada anteriormente al desaparecer alguna dimensión (por ejemplo,

competitividad y casi *cooperación* y *ambiente físico* —tabla 4—) lo cual fue imprescindible para mejorar la baja fiabilidad inicial. En secundaria, el criterio sustantivo se cumple en cuanto a que todas las dimensiones quedan representadas en la versión reducida, con la salvedad de haber incluido el ítem 17 de la dimensión *interés* y *preocupación* que, aunque no cumplía uno de los criterios de fiabilidad —tabla 5—, no disminuía el alfa general del micro-instrumento —tabla 2—.

Discusión

La primera fase de la R-AVACO consistente en los pasos a), b) y c) ha tenido como objetivo reducir la escala base a fin de configurar una versión menor para las dos muestras con las que se ha trabajado: sexto de primaria y cuarto de secundaria. En esta fase los análisis se han centrado en el estudio pormenorizado de los ítems, decidiendo sobre su calidad para formar parte de la misma. Cinco son los ítems comunes que han sido seleccionados en las dos muestras: (20) *me siento orgulloso de mi trabajo en clase*, (25) *la mayoría de los compañeros estamos satisfechos con lo que aprendemos en clase*, (26) *al finalizar el curso voy a tener un buen nivel para continuar en los cursos siguientes*, (31) *la mayor parte de mis compañeros deja trabajar a los demás* y (34) *la distribución de la clase permite trabajar a gusto*. De ellos, el reactivo que ha mostrado mayor calidad es el ítem 25: además de superar todos los indicadores de reducción, ha presentado coeficientes de correlación ítem-total elevados (tabla 2) y un rango también menor al de los restantes reactivos (tablas 6 y 7). Por muestras, considerando los resultados combinados de los análisis de conglomerados y de fiabilidad, han destacado el ítem 31 en primaria y el 26 en secundaria.

La segunda fase de la R-AVACO ha consistido en los pasos d) y e), realizando la validación del conjunto de las versiones reducidas. Tal y

como indica el procedimiento R-AVACO, podemos ahora valorar los dos micro-instrumentos en relación a las características de eficacia, eficiencia y funcionalidad (recordemos la figura 1).

La máxima *eficiencia* es posible juzgarla a partir de haber maximizado tanto la correlación entre las puntuaciones totales como las asociaciones entre las clasificaciones de las unidades muestrales. En primaria los resultados han aconsejado no reducir más el instrumento porque sería posible que los criterios de validación que todavía se cumplen dejaran de hacerlo. El micro-instrumento obtenido es poco robusto en cuanto a la presencia de *missing* que, aun no siendo superiores al 5% en ningún ítem, no han podido confirmarse como aleatorios. En secundaria, la eficiencia del micro-instrumento ha resultado suficiente según los criterios aplicados en el proceso de validación, lo que permite confiar en la selección de elementos efectuada. En relación a la presencia de valores ausentes, el micro-instrumento de secundaria ha resultado francamente sólido.

En cuanto a la *funcionalidad*, sustantivamente se ha perdido la dimensión de *competitividad* en la muestra de primaria, lo que sugiere quizá una revisión de la adecuación de los ítems empleados según las características de esta muestra. En secundaria, todos los ítems de la dimensión *interés* y *preocupación* carecen de las características suficientes para ser seleccionados. Al objeto de no perder otras dimensiones, se ha optado por mantener algunos elementos que no cumplen el criterio de presencia de valores ausentes (por ejemplo, en primaria, los reactivos 30 y 31 de *cooperación* y el ítem 34 de *ambiente físico*). Con todo, consideramos que el proceso seguido ha sido escrupuloso, salvando la única excepción del ítem 17 (*la mayor parte de mis compañeros quiere aprender*) que, sin haber superado uno de los indicadores de reducción, ha quedado incorporado para no perder la dimensión de *interés* y *preocupación* en secundaria.

Respecto a la *eficacia*, la equivalencia entre la escala base y la versión reducida ha sido comprobada en ambas muestras con las pruebas *Ji* cuadrado tanto globalmente como por centros. En cuanto a las asociaciones, han resultado algo bajas en dos de los cinco centros, también en las dos muestras.

Realizando una valoración conjunta de todo el proceso, es destacable la diferencia de resultados entre primaria y secundaria, cuando en ambos grupos se ha partido de una misma escala base. Esto subraya la importancia de la sensibilidad de los instrumentos para poder adaptarse a las características de las muestras con las que se trabaja. Aunque se trataba de medir el mismo constructo —clima de aprendizaje—, ha quedado de manifiesto que el comportamiento de los ítems y del conjunto de la escala es distinto dependiendo de una muestra u otra. Es evidente que los cuatro años de distancia en la edad de los alumnos comportan diferencias a la hora de percibir de un modo u otro los elementos de la escala. Esta podría ser una explicación de la pérdida de la dimensión de *competitividad* en la muestra de primaria.

Estas diferencias en las muestras se han visto también reflejadas en la presencia de observaciones ausentes. Son los alumnos de primaria los que tienden a dejar mayor número de preguntas sin contestar y ello ha influido notablemente en el análisis de los ítems afectados. Cabría pensar en una posible dificultad de los ítems para este tipo de alumnos, tal vez debido a su redacción o quizá por su contenido. No podemos establecer conclusiones al respecto. Lo que está claro es que la presencia de valores ausentes condiciona de forma notable cualesquiera análisis estadísticos que deseen hacerse con posterioridad, bien sobre el estudio de las características del instrumento (como aquí se ha hecho), bien respecto al contenido de las respuestas de los encuestados. La solución parece apuntar en la misma línea expuesta en el párrafo anterior: debe procurarse el diseño de instrumentos adaptados a las muestras de estos

niveles de enseñanza sin dar por hecho una equivalencia entre ellas.

Ahora bien, en este punto debemos matizar un aspecto que sin duda es preciso mejorar. Para poder extraer las conclusiones señaladas sería conveniente trabajar al menos con dos grupos de cada nivel de edad. Utilizar únicamente una muestra de cada nivel escolar no parece concluyente. Un posible modo de mejorar el estudio sería haciendo particiones (al azar) en las muestras originales y comprobar si los resultados siguen siendo los mismos.

En cuanto a los criterios de diseño que mencionamos en la introducción de este trabajo, el procedimiento de reducción R-AVACO ha demostrado ser francamente útil para cumplir todos ellos: se han obtenido dos micro-instrumentos válidos y fiables, con ítems que discriminan bien a los sujetos, con significación sustantiva y suficientemente parsimoniosos. Todo ello abunda en la confianza de poder diseñar micro-escalas de contexto con garantías de calidad, lo que sin duda revierte en la mejora de la evaluación de un sistema educativo.

Notas

¹ Este trabajo se desarrolló en el contexto del proyecto de investigación *Análisis de Variables de Contexto: Diseño de cuestionarios de contexto para la evaluación de sistemas educativos* (Proyecto AVACO, I+D+I, 2006-2008. Código SEJ 2005-05 923). Investigador principal: Jesús Jornet Meliá, del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universitat de València.

² Estos procesos se describen con detalle en el artículo de Pérez-Carbonell, Ramos y López-González (2009), efectuados en ese caso con el constructo *clima social aula*.

³ La prueba MCAR —inicialmente introducida por Rubin (1977)— se basa en el supuesto de que los datos siguen un patrón completamente aleatorio (*missing completely at random*). Este supuesto es asumido en la mayoría de las situaciones en las que se elige imputar observaciones ausentes (Medina y Galván, 2007). Lo habitual en la práctica es que no se satisfaga el criterio, ya que la falta de respuestas suele deberse más bien a características de las personas encuestadas. La prueba realiza un test con J_i cuadrado entre las dos condiciones (datos completamente aleatorios —que incluyen los valores imputados— y datos observados). Para aceptar el azar no debe haber significación estadística.

⁴ Se sigue este criterio por indicación de la reducción R-AVACO (Jornet *et al.*, 2012).

Referencias bibliográficas

- BIENCINTO, C., GONZÁLEZ-BARBERA, C., GARCÍA-GARCÍA, M., SÁNCHEZ-DELGADO, P. y MADRID, D. (2009). Diseño y propiedades psicométricas del AVACO-EVADIE. Cuestionario para la evaluación de la atención a la diversidad como dimensión educativa en las instituciones escolares, *Relieve*, 15 (1).
- JORNET, J. M., GONZÁLEZ-SUCH, J. y PERALES, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos, *Bordón*, 64 (2), 89-110.
- LITTLE, R. J. A. y RUBIN, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.
- LÓPEZ-GONZÁLEZ, E., PÉREZ-CARBONELL, A. y RAMOS-SANTANA, G. (2011). Modelos complementarios al Análisis Factorial en la construcción de escalas ordinales: un ejemplo aplicado a la medida del clima social aula, *Revista de Educación*, 354, 369-397.
- MEDINA, F. y GALVÁN, M. (2007). *Imputación de datos: teoría y práctica*. Santiago de Chile: CEPAL (Naciones Unidas).
- PÉREZ-CARBONELL, A., RAMOS-SANTANA, G. y LÓPEZ-GONZÁLEZ, E. (2009). Diseño y análisis de una escala para la valoración de la variable clima social aula en alumnos de educación primaria y secundaria, *Revista de Educación*, 350, 221-252.

RUBIN, D. B. (1977). Formalizing subjective notions about the effect of non-respondents in sample surveys, *Journal of the American Statistical Association*, 72.

Abstract

Designing a micro instrument for the measurement of learning climate useful to develop context questionnaires

We work with an initial context questionnaire to measure the learning climate construct in two school samples corresponding to the 6th level of primary education and the 4th level of secondary education. The objective was to design a parsimonious micro-instrument which would be valid, reliable, and with enough capacity to discriminate among subjects and with substantive significance. For this reason we designed a shorter version of the original scale, applying the R-AVACO reduction procedure. In the first stage of this procedure those key items with higher quality for the proposed objectives are selected. Analysis of missing values, exploratory analysis, cluster analysis and reliability analysis are used. In a second stage a global validation of the micro-instrument is performed. As a result, two shorter versions are obtained and fitted to the small samples, noting the importance of missing values and merging the results of several statistical analyses carried out in the preceding five steps on R-AVACO.

Key words: *Evaluation of education systems, Learning climate, Context questionnaires, Reliability, Validity.*

Résumé

La conception d'un micro-instrument pour mesurer le climat d'apprentissage dans les questionnaires contextuels

Cet article travail à partir d'une échelle de base comme questionnaire de contexte qui mesure le climat d'apprentissage. Nous cherchons à simplifier la mesure du construit dans l'évaluation des systèmes éducatifs. L'étude a été menée auprès de deux échantillons d'élèves de 6^{ème} cours d'Éducation Primaire et de 4^{ème} cours d'Éducation Secondaire Obligatoire, respectivement. L'objectif est de concevoir un micro-instrument parcimonieux, fiable, valide et avec une suffisante capacité pour différencier les élèves. Ceci est déterminé avec une version diminuée de l'échelle de base en utilisant la procédure de réduction R-AVACO. La conception est effectuée en deux phases. Dans un premier moment, les items de meilleure qualité et avec le plus grand pouvoir de prédiction sont sélectionnés, employant diverses techniques statistiques: analyses exploratoires et études des observations manquantes, analyses des conglomerats et analyses de fiabilité. Dans une deuxième phase, nous estimons les valeurs totales, nous explorons la version réduite de l'échelle de base et nous validons le micro-instrument. Comme résultat, nous avons obtenus deux versions réduites de l'échelle de base initiale, adaptées aux échantillons avec lesquels nous travaillons. Ces micro-instruments son finalement évalués selon des critères d'efficacité, de fonctionnalité et d'efficacité.

Mots clés: *Évaluation des systèmes éducatifs, Climat d'apprentissage, Questionnaire de contexte, Fiabilité, Validité.*

Perfil profesional de los autores

Emelina López-González

Profesora titular de Psicoestadística y de Métodos de Investigación en Educación. Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación, Universidad de Valencia.

Sus principales líneas de investigación son: desarrollo y aplicación de los modelos lineales generalizados y modelos multivariados para el análisis de datos categóricos en el ámbito de las Ciencias Sociales.

Correo electrónico de contacto: emelina.lopez@uv.es

Javier Tourón

Catedrático de Métodos de Investigación y Diagnóstico en Educación en el Departamento de Educación de la Universidad de Navarra. Past-President del European Council for High Ability. Miembro del National Advisory Board del Center for Talented Youth de la Universidad Johns Hopkins (2003-2011). Fundador y director de CTY España (2001-2011). Sus líneas de investigación se centran en el estudio y desarrollo de procedimientos de identificación e intervención educativa con alumnos de alta capacidad, en el desarrollo y análisis de instrumentos de medida en educación, así como en la evaluación de sistemas educativos.

Correo electrónico de contacto: jtouren@unav.es

Francisco Javier Tejedor

Catedrático en la Facultad de Educación de la Universidad de Salamanca. Sus líneas principales de investigación son: evaluación del rendimiento de los estudiantes universitarios, competencias TIC de los estudiantes universitarios, evaluación de procesos de innovación con TIC en distintos contextos educativos y desarrollo profesional docente.

Correo electrónico de contacto: tejedor@usal.es