



CIENCIAS HUMANAS Y SOCIALES

**ESTUDIO SOBRE UN EXAMEN DE TIPO
TEST BASADO EN EL GRADO DE
CONFIANZA**

Autor: Carlos Valencia Rodríguez

Directores: Belen Urosa Sanz, Pedro Morales Vallejo

Madrid

Diciembre de 2015

ÍNDICE

ÍNDICE DE TABLAS	11
ÍNDICE DE GRÁFICOS	17
1. CAPÍTULO I: INTRODUCCIÓN	19
2. CAPÍTULO II: MARCO CONCEPTUAL Y ESTADO DE LA CUESTIÓN	25
2.1. FINALIDAD DE LA EVALUACIÓN DEL ESTUDIANTE	26
2.1.1. Procedimientos de evaluación mediante pruebas objetivas de elección múltiple	31
2.1.1.1. Otras críticas a los exámenes de tipo test	45
2.1.1.2. No valoración del conocimiento parcial en los test	46
2.1.1.3. Posibilidad de acierto por azar	49
2.1.2. Corrección de la adivinación en los exámenes de tipo test	54
2.1.3. Inconvenientes de los sistemas de corrección que penalizan las respuestas erróneas	58
2.1.4. Alternativas al método tradicional de corrección del azar en los exámenes de tipo test	61
2.2. LA CONFIANZA COMO INDICADOR DE APRENDIZAJE AUTORREGULADO	63
2.2.1. Aprendizaje autorregulado	63
2.2.1.1 Automonitarización/Metacognición /	69
2.2.1.2. Autoeficacia	76
2.2.1.3. Confianza académica	79
2.2.2.- Los exámenes y la evaluación de la confianza.....	86

2.3. ENFOQUES DE APRENDIZAJE Y EVALUACIÓN	90
2.3.1. Enfoques de aprendizaje	90
2.3.1.1. Enfoques superficial y profundo	92
2.3.1.2. Enfoques de aprendizaje y metacognición	94
2.4. EXAMEN DE TIPO TEST BASADO EN LA CONFIANZA (TBC)	96
2.4.1. Justificación de este tipo de evaluación	96
2.4.2. Descripción general del procedimiento	97
2.4.3. Objetivos y ventajas del procedimiento de evaluación basado en la confianza	98
2.4.4. El TBC como procedimiento para distinguir el conocimiento seguro del inseguro	101
2.4.5. Importancia en las evaluaciones de la confianza del estudiante	104
2.4.6. El factor de la sinceridad del alumno en el momento de escoger su opción de seguridad	106
2.4.6.1. Importancia de un sistema de examen que valore la confianza además de los conocimientos	107
2.4.7. Modelos de puntuación de test basados en la confianza	108
2.4.7.1. Un poco de historia	108
2.4.7.2. Tabla explicativa de las características de los sistemas de TBC ...	109
2.4.7.3. Modelo de Ahlgren (1970)	110
2.4.7.4. Modelo de Hassmen y Hunt	111
2.4.7.5. Modelos de Davies	112
2.4.7.6. Modelo de Sainsbury y Walker	114
2.4.7.7. Modelo de Gardner-Medwin	115
2.4.7.8. Modificación del sistema de Gardner-Medwin	118
2.4.8. Validez y fiabilidad del procedimiento de corrección basado en	

la confianza	120
2.4.8.1. Introducción al tema de la validez y fiabilidad del TBC	120
2.5. RESUMEN SOBRE EL ESTADO DE LA CUESTIÓN.....	121

3. CAPÍTULO III: METODOLOGÍA DEL ESTUDIO

3.1. OBJETIVO DE LA INVESTIGACIÓN	125
3.2. MUESTRAS UTILIZADAS EN EL ESTUDIO	128
3.2.1. Participantes incluidos en el estudio de validez	128
3.2.2. Participantes incluidos en el estudio de fiabilidad	131
3.2.3. Participantes y grupos en las encuestas	133
3.3. INSTRUMENTOS	136
3.3.1. Estudio preliminar	136
3.3.1.1. Formato de los exámenes de TBC utilizados por nosotros ...	137
3.3.1.2. Materias en las que se realiza el estudio del TBC.....	140
3.3.2. Instrumentos: Encuestas	145
3.3.2.1. Ítems de la encuesta	147
3.3.3. Procedimientos	154
3.3.3.1. Procedimiento para el análisis de la validez del TBC en comparación con el sistema de test convencional	154
3.3.3.2. Procedimiento para el análisis de la consistencia interna del TBC.....	156
3.3.3.3. Procedimiento seguido con la encuesta	156
3.3.3.4. Procedimiento para el cálculo del nivel de confianza de los estudiantes	158

3.3.3.5 – Puntuación en confianza (PC)	158
3.3.3.6. Índice de confianza (IC) y <i>PC–nota por aciertos</i>	164
3.4. OBJETIVOS E HIPÓTESIS DEL ESTUDIO	167
3.4.1. Objetivos	167
3.4.2. Hipótesis que queremos contrastar	169
3.5. RECOGIDA DE DATOS	170
3.5.1. Recogida de datos del estudio de consistencia interna	170
3.5.2. Recogida de datos de la encuesta	171

4. CAPÍTULO IV: RESULTADOS

4.1. ESTUDIO PRELIMINAR PARA ESTABLECER LA VALIDEZ Y FIABILIDAD DEL TBC	
4.1.1. Estudio de la validez del examen de tipo test basado en la confianza (TBC)	177
4.1.1.1. Adecuación del nivel de seguridad a los conocimientos reales del alumno	177
4.1.1.2. Validez del TBC en comparación con la validez del test por Aciertos.....	180
4.1.2. Estudio de la consistencia interna de los exámenes de tipo test, comparando el TBC con el test por aciertos	181
4.2. RESULTADOS DE LA ENCUESTA	184
4.2.1. Resultados de la encuesta pasada a los alumnos en febrero-marzo de 2013 y 2014	186
4.2.1.1. Encuesta de la valoración de los alumnos de los diferentes modelos de examen (ítems 2, 3 y 4)	187

4.2.1.2. Influencia de los diversos tipos de examen sobre los hábitos de estudio de los alumnos	194
4.2.1.3. Ítems que valoran la <i>validez percibida por el alumno</i> de cada tipo de examen. Descripción	207
4.2.1.4. Ítems que valoran la <i>validez percibida por el alumno</i> de cada tipo de examen. Análisis del agrupamiento y correlaciones ...	212
4.2.1.5. Agrupamiento de las preguntas que miden la <i>eficacia docente percibida por el alumno</i> de cada tipo de examen	232
4.2.1.6. Relación entre el sistema de examen y la generación de estrés al alumno	248
4.2.2. Relación entre el sistema de examen y los enfoques de aprendizaje, profundos y superficiales	253
4.2.2.1. Correlaciones de los enfoques profundos de aprendizaje ...	256
4.2.2.2. Correlaciones entre el rendimiento académico y los enfoques de aprendizaje superficiales	260
4.2.2.3. Diferencias en las correlaciones «enfoque de aprendizaje– resultados académicos» entre el TBC y el test <i>por aciertos</i> ...	262
4.2.2.4. Ítem «TBC–aciertos»	263
4.2.2.5. Correlaciones del ítem «TBC–aciertos»	264
4.2.2.6. Correlaciones entre la habilidad en el manejo del TBC y los enfoques de aprendizaje del alumno	266
4.2.2.7. Correlaciones entre los enfoques superficiales de aprendizaje y el índice <i>TBC–aciertos</i>	268
4.2.3. Estudio de los resultados obtenidos en los ítems que valoran la autorregulación del alumno	269

4.2.3.1. Correlaciones de los ítems de autorregulación con los resultados académicos	271
4.2.3.2. Correlaciones de los ítems de autorregulación con los enfoques de aprendizaje	274
4.2.3.3. Correlaciones de otros ítems con la autorregulación	275
4.3. ESTUDIO DE LA INFLUENCIA DEL GÉNERO DE LOS ESTUDIANTES EN LOS RESULTADOS DE LA ENCUESTA	279
4.4. ESTUDIO FACTORIAL DE LA ENCUESTA	282
4.4.1. Escala R-SPQ–2F (de Biggs)	283
4.4.1.1. Primer componente	284
4.4.1.2. Segundo componente	286
4.4.1.3. Tercer componente	287
4.4.1.4. Cuarto componente	289
4.4.1.5. Quinto componente	291
4.4.2. Análisis factorial de la escala RSPQ–2F con dos agrupamientos...	292
4.4.3. Análisis factorial de las preguntas sobre <i>Autorregulación</i>	293
4.4.4. Análisis factorial de los demás ítems de la encuesta	295
4.4.4.1. Componentes que miden la « <i>eficacia docente</i> » de cada tipo de examen	297
4.4.4.2. Otros agrupamientos que se obtienen del análisis factorial..	300
4.5. ESTUDIO DE LOS PARÁMETROS QUE MIDEN LA AUTOCONFIANZA MEDIA DEL ALUMNO	304

4.5.1. Estudio de la <i>Puntuación en Confianza</i> y del <i>Índice de Confianza</i> ...	304
4.5.1.1. Correlación entre la PC y el resultado académico	305
4.5.1.2. Resultados obtenidos en la PC, en el índice de confianza (IC) y en <i>PC- aciertos</i>	306
4.5.2. Relación entre la autoconfianza y los resultados del examen	308
4.5.3 Relación entre la autoconfianza y el género de los estudiantes	311
5. CAPÍTULO V: CONCLUSIONES.....	315
5.1. VALIDEZ DEL TBC	316
5.2. CONSISTENCIA INTERNA DEL TBC	317
5.3. ACEPTACIÓN DEL TBC	318
5.4. MODIFICACIÓN EN LOS HÁBITOS DE ESTUDIO	320
5.5. OPINIÓN DE LOS ALUMNOS SOBRE LAS CUALIDADES DEL TBC EN COMPARACIÓN CON LOS DOS OTROS TIPOS DE EXAMEN	323
5.6. ESTUDIO DE LOS ENFOQUES DE APRENDIZAJE DE LOS ALUMNOS Y SU CORRELACIÓN CON LA OPINIÓN SOBRE LOS DIVERSOS TIPOS DE EXAMEN Y CON SUS RESULTADOS CON EL TBC	333
5.7. CONCLUSIONES DEL ESTUDIO DE LOS ÍTEMS DE AUTORREGULACIÓN Y SUS CORRELACIONES	347
5.8. INFLUENCIA DEL GÉNERO DE LOS ESTUDIANTES EN LOS RESULTADOS OBTENIDOS CON EL TBC	350
5.9. CONCLUSIONES DEL ESTUDIO DE LA AUTOCONFIANZA DE LOS ALUMNOS	351
5.10. RELACIÓN ENTRE GÉNERO Y AUTOCONFIANZA	360

5.11. CONCLUSIONES RELEVANTES RESPECTO A LOS OTROS DOS SISTEMAS DE EXAMEN	362
----------------------------------------------------------------------------------------	------------

6. CAPÍTULO VI: DISCUSIÓN

6.1. CARACTERÍSTICAS DE UN BUEN SISTEMA DE EVALUACIÓN	367
6.2. VALIDEZ	370
6.3. FIABILIDAD	372
6.4. ACEPTACIÓN DEL TBC POR EL ESTUDIANTE	375
6.5. EXÁMENES Y ESTRÉS DEL ESTUDIANTE	376
6.6. INFLUENCIA DEL TBC EN LOS HÁBITOS DE ESTUDIO DEL ALUMNO	377
6.7. OPINIONES DE LOS ALUMNOS SOBRE LA VALIDEZ Y LA EFICACIA DOCENTE DEL TBC	378
6.8. TBC Y GÉNERO	379
6.9. TBC Y ENFOQUES DE APRENDIZAJE	383
6.10. TBC Y AUTORREGULACIÓN	389
6.11. MEDICIÓN DE LA CONFIANZA ACADÉMICA CON EL TBC	391
6.12. CONCLUSIÓN FINAL	397

7. BIBLIOGRAFÍA	401
------------------------------	------------

8. ANEXOS.....	419
-----------------------	------------

ANEXO A	419
----------------------	------------

ANEXO B	425
ANEXO C	426
ANEXO D	427
ANEXO E	428
ANEXO F	435
ANEXO G	445
ANEXO H	459

ÍNDICE DE TABLAS

TABLAS DEL CAPÍTULO 2	PÁGINA
2.1 – Modificación por Anderson y Krathwohl del diagrama de Bloom	30
2.2 – Taxonomía revisada de las guías para componer los ítems de un test de elección múltiple.	35-37
2.3 – Diferencia entre autoeficacia, autoconcepto y autoconfianza	80
2.4 – Test ABC	82-83
2.5 - Modelo de comportamiento de aprendizaje	99
2.6 - Situaciones que se presentan al responder un alumno a una pregunta de test	102
2.7 - Cuadrante de posibilidades al responder a una pregunta	102
2.8 - Modelo de tabla utilizada para analizar los sistemas de puntuación	110
2.9 - Modelo de TBC de Ahlgren (1970)	111
2.10 - Modelo de TBC de Hassmen y Hunt (1994) y Khan (2001)	112
2.11 - Modelo de TBC, A de Davies	113
2.12 - Modelo de TBC, B de Davies	113
2.13 - Modelo de TBC, C de Davies	113
2.14 - Modelo de TBC de Sainsbury y Walker (2008), propuesta 1	114
2.15 - Modelo de TBC de Sainsbury y Walker (2008), propuesta 2	115
2.16 - Modelo de TBC de Gardner–Medwin para 2 opciones de respuesta	117
2.17 - Modelo de TBC de Gardner–Medwin para 3 o más opciones de respuesta	117
2.18 - Tabla de puntuación utilizada por nosotros desde el año 2012	119

TABLAS DEL CAPÍTULO 3	PÁGINA
3.1 - Individuos y grupos que participan en el estudio de validez	131
3.2 - Participantes en el estudio de fiabilidad y número de preguntas en cada examen	131
3.3 - Número de participantes en cada grupo y año en la encuesta	134
3.4 - Modificación del porcentaje de respuestas correctas en cada nivel de seguridad en dos exámenes sucesivos.	136
3.5 - Individuos y grupos que participan en el estudio de validez	141
3.6 - Grupo, nombre de la materia y número de sujetos de cada grupo, estudio de fiabilidad.	141
3.7 - Tabla de puntuaciones utilizada en nuestros exámenes de TBC	143
3.8 - Tabla de puntuaciones del <i>test por aciertos</i>	143
3.9 - Características de los datos recogidos en la encuesta	147
3.10 - Variables de la encuesta	148-149
3.11 - Puntuación asignada a los tres niveles de confianza	159
3.12 - Puntuación media obtenida por los alumnos con cada nivel de confianza	159

3.13 - Puntuación de cada nivel de confianza, corrigiendo el mínimo a «0»	159
3.14 - Puntuación de cada nivel de confianza, igualando el máximo a «10»	160
3.15 - Relación entre la nota media (por aciertos) según el nivel de confianza y puntuación en confianza.	162
3.15 - Comparación de las notas medias en el test por aciertos y examen de preguntas abiertas y el nivel medio de confianza	162
3.16 - Grupos incluidos en la encuesta de nuestro trabajo	172

TABLAS DEL CAPÍTULO 4	PÁGINA
4.1 - Puntuación media de nuestros estudiantes con tres sistemas de examen	178
4.2 - Porcentaje de aciertos con cada nivel de seguridad	179
4.3 - Correlación de la nota de cada tipo de test con la nota obtenida en el examen de preguntas abiertas	181
4.4 - Puntuaciones según el nivel de confianza y el test por aciertos	182
4.5 - Fiabilidad de los exámenes según el sistema de corrección	183
4.6 - Pregunta sobre la aceptación de los tres tipos de examen	187, 318
4.7 - Significación estadística de las diferencias en la aceptación media de cada examen	188
4.8 - Aceptación entre los encuestados de TBC	189
4.9 - Correlaciones de los ítems de preferencia por exámenes (2, 3, 4)	190-191
4.10 - Influencia del tipo de examen sobre los hábitos de estudio del alumnos	194
4.11 - Tiempo dedicado al estudio con cada tipo de examen	195
4.12 - Significación estadística de las diferencias	196
4.13 - Tiempo dedicado al estudio con cada tipo de examen	197
4.14 - Correlaciones de los ítems que valoran el tiempo de estudio con los demás	198
4.15 - Correlaciones de los ítems de «tiempo de estudio» entre sí	199
4.16 - Medida en la que cada examen obliga al alumno a comprender lo que aprende	200
4.17 - Significación estadística de las diferencias en la media con los tres exámenes	200
4.18 - Tabla de correlaciones de los ítems 8, 9 y 10 con las preguntas que interrogan sobre la <i>validez autopercebida</i> del TBC	202
4.19 - Correlaciones de la pregunta sobre el esfuerzo de comprensión que cada tipo de examen exige con la escala de enfoques de aprendizaje y la de autorregulación	204
4.20 - Datos estadísticos de la pregunta sobre el examen que obliga a memorizar más (ítems 11, 12 y 13)	205
4.21 - Significación estadística de las diferencias en las medias con los tres exámenes	205
4.22 - Capacidad de cada examen para informar al profesor de los conocimientos del alumno	209
4.23 - Significación estadística de las diferencias en las medias de los ítems 15, 16 y 17.	210
4.24 - Respuesta a la pregunta sobre la influencia de cada tipo de examen en la autoevaluación del estudiante	210

4.25 - Significación estadística de las diferencias en la media de los tres tipos de examen a la pregunta anterior	211
4.26 - Respuesta a la pregunta sobre la influencia de cada examen en «aprender más»	211
4.27 - Significación estadística de las diferencias en la media de la tabla 4.23	212
4.28 - Correlación entre los ítems que miden la <i>validez autopercebida</i> del test <i>por aciertos</i>	213
4.29 - Correlaciones entre los ítems que miden la <i>validez autopercebida</i> del TBC	214
4.30 - Correlación entre los ítems que miden la <i>validez</i> de las <i>preguntas abiertas</i>	215
4.31 - Correlaciones de los tres ítems que miden la <i>validez</i> del test <i>por aciertos</i>	216
4.32 - Correlaciones de los tres ítems que miden la <i>validez</i> del TBC	216
4.33 - Correlaciones de los tres ítems que miden la <i>validez</i> del examen de <i>preguntas abiertas</i>	216
4.34 - Media de las puntuaciones del agrupamiento <i>validez autopercebida</i> de cada examen	217
4.35 - Correlaciones del agrupamiento que expresa la «validez autopercebida» de cada examen	219
4.36 - Correlaciones de los ítems de satisfacción con los exámenes con los enfoques de aprendizaje y la autorregulación - Correlaciones de los ítems de de apreciación del esfuerzo de comprensión con los exámenes con los enfoques de aprendizaje y la autorregulación - Correlaciones de los agrupamientos de <i>validez autopercebida</i> con los enfoques de aprendizaje y la autorregulación	221-222
4.37 - Otras correlaciones de los agrupamientos de «validez autopercebida» de los exámenes	223
4.38 - Correlación con ítems que valoran la necesidad de comprender bien los conceptos estudiados	224
4.39 - Correlación del agrupamiento que valora la «validez» de cada examen con ítems que valoran el aprecio del alumno con cada tipo de examen	225
4.40 - Correlaciones con los ítems que valoran el tiempo dedicado al estudio con cada examen con los agrupamientos	226
4.41 - Correlaciones con los ítems que valoran esfuerzo de memorización con cada examen con los agrupamientos	227
4.42 - Tabla que muestra el esfuerzo de memorización que exige cada examen	228
4.43 - Correlaciones de los agrupamientos con el acierto al escoger la opción de seguridad	229
4.44 - Correlaciones entre los exámenes preferidos por los alumnos y su opinión sobre la «validez» de cada tipo de examen	229
4.45 - Correlación entre el nivel de estrés que genera cada examen en el alumno y su opinión sobre la «validez» de cada examen	230
4.46 - Correlaciones de los grupos que valoran la opinión del alumno sobre la «validez autopercebida» de los exámenes	231
4.47 - Correlación entre la «validez» de cada tipo de examen y el tiempo dedicado al estudio y el esfuerzo de comprensión	233
4.48 - Agrupamientos que definen la <i>eficacia docente</i> de cada tipo de examen	234, 297
4.49 - Correlaciones internas de los ítems de <i>eficacia docente</i> del test <i>por aciertos</i>	235

4.50 - <i>Eficacia docente</i> del test por aciertos: tabla de correlaciones de cada ítem con el agrupamiento	235
4.51 - Correlaciones internas de los ítems de <i>eficacia docente</i> del TBC	236
4.52 - <i>Eficacia docente</i> del TBC. Tabla de correlaciones de cada ítem con los demás del agrupamiento o <i>peso</i> cada ítem	236
4.53 - Correlaciones internas de los ítems de <i>eficacia docente</i> del <i>examen de preguntas abiertas</i>	237
4.54 - <i>Eficacia docente</i> del examen de preguntas abiertas: la tabla de correlaciones de cada ítem con el agrupamiento	237
4.55 - Correlación entre los tres agrupamientos, correspondientes a cada tipo de examen, que miden la <i>eficacia docente</i>	238
4.56 - Puntuación media en la encuesta a cada uno de los agrupamientos de <i>eficacia docente</i>	240
4.57 - Significación estadística de las diferencias en la media entre los tres exámenes	240
4.58 - Correlación de <i>eficacia docente</i> de cada examen con enfoques de aprendizaje y autorregulación	242, 332
4.59 - Correlación de los agrupamientos <i>validez</i> y <i>eficacia</i> de la opinión sobre el TBC con enfoques de aprendizaje y autorregulación	243
4.60 - Correlaciones del ítem 29 con la opinión favorable a cada tipo de examen	244
4.61 - Correlaciones del ítem 31 con la <i>eficacia docente</i> percibida por el alumno en cada tipo de examen	245
4.62 - Correlación del agrupamiento de opinión favorable sobre la eficacia de cada tipo de examen y el agrupamiento de autorregulación	246
4.63 - Otras correlaciones de los agrupamientos de eficacia de cada examen	247
4.64 - Puntuaciones medias de los estudiantes al estrés que perciben con cada tipo de examen	249, 320
4.65 - Significación estadística de las diferencias	249
4.66 - Tabla de correlaciones de los ítems 26, 27 y 28 con los exámenes	250
4.67 - Correlación estrés – género	250
4.68 - Estrés de cada género con los tres tipos de examen. Ítem 26	251
4.69 - Estrés de cada género con los tres tipos de examen. Ítem 27	251
4.70 - Estrés de cada género con los tres tipos de examen. Ítem 28	251
4.71 - Correlaciones de los ítems que miden estrés con cada examen y los otros ítems de la encuesta	252
4.72 - Tabla de estadísticos básicos de los ítems que miden enfoques profundos	254
4.73 - Tabla de estadísticos básicos de los ítems que miden enfoques superficiales	254
4.74 - Comparación de los enfoques profundos con los superficiales (1)	255
4.75 - Comparación de los enfoques profundos con los superficiales (2)	255
4.76 - Comparación de los enfoques profundos con los superficiales (3)	255
4.77 - Correlación motivo profundo–resultados en el examen	256
4.78 - Correlación estrategia profunda–resultados en el examen	257
4.79 - Correlación entre los ítems de enfoque profundo y los resultados en los exámenes	257
4.80 - Correlaciones entre los ítems de enfoque profundo y los resultados en los exámenes	258

4.81 - Correlaciones entre los ítems de motivo profundo y los resultados en los exámenes	258
4.82 - Correlaciones entre los ítems de estrategia profunda y los resultados en los exámenes	259
4.83 - Correlaciones entre los ítems de enfoque superficial y los resultados en los exámenes	260
4.84 - Correlaciones entre los ítems de motivo superficial y los resultados en los exámenes	260
4.85 - Correlaciones entre los ítems de estrategia superficial y los resultados en los exámenes	261
4.86 - Estadísticos de los resultados en el examen	263
4.87 - Correlación entre la habilidad en el manejo del TBC y las notas por aciertos	264
4.88 - Correlaciones del TBC– <i>aciertos</i> con los otros ítems	265
4.89 - Correlaciones entre TBC– <i>aciertos</i> y factor «enfoque profundo»	266
4.90 - Correlaciones entre TBC– <i>aciertos</i> y estrategia profunda	267
4.91 - Correlaciones entre TBC– <i>aciertos</i> y motivo profundo	267
4.92 - Correlaciones entre TBC– <i>aciertos</i> y enfoques superficiales	268
4.93 - Correlaciones entre los ítems de autorregulación	270
4.94 - Correlaciones de cada ítem de autorregulación con la suma del resto de ítems	270
4.95 - Correlación entre la autorregulación y las puntuaciones en el examen	271
4.96 - Correlaciones de las puntuaciones del examen con cada ítem de autorregulación	272
4.97 - Otras correlaciones de la autorregulación	273
4.98 - Otras correlaciones significativas de los cinco factores incluyendo autorregulación	274
4.99 - Correlación autorregulación – enfoques de aprendizaje	274
4.100 - Correlación autorregulación – motivos y estrategias de aprendizaje	274
4.101 - Correlaciones de la autorregulación con ítems individuales	275
4.102 - Correlaciones de la autorregulación con los agrupamientos que miden la «validez» de cada tipo de examen	276
4.103 - Correlaciones de la autorregulación con los agrupamientos que miden la «eficacia docente» de cada tipo de examen y el esfuerzo que exige al alumno	276
4.104 - Correlación autorregulación con ítems de «opinión positiva» sobre el TBC	277
4.105 - Correlación autorregulación con ítems sobre el tipo de aprendizaje que induce cada uno de los tres tipos de examen	278
4.106 - Correlaciones entre el género de alumno y los ítems de la encuesta	279
4.107 - Motivos profundos (suma ítems 29, 33, 37, 41 y 45) en cada género	280
4.108 - Estrategia profunda (suma de ítems 30, 34, 38, 42 y 46) en cada género	280
4.109 - Motivo superficial (suma de ítems 31, 35, 39, 43 y 47) en cada género	281
4.110 - Estrategia superficial (suma ítems 32, 36, 40, 44 y 48) en cada género	281
4.111 - Estadísticos de los resultados académicos separados por género	281
4.112 - Componentes válidos del análisis factorial de la encuesta	283
4.113 - Matriz de componentes rotados, SPQ–2F	284
4.114 - Tabla de correlaciones del primer componente	285
4.115 - Tabla de correlaciones internas del segundo componente	286
4.116 - Tercer componente (ítems 36, 39, 30 y 38)	287

4.117 - Cuadro de correlaciones internas del cuarto componente	289
4.118 - Cuadro de correlaciones internas del quinto componente	291
4.119 - Agrupamientos y correlaciones de la encuesta SPQ-2F con dos componentes	292
4.120 - Análisis factorial de los ítems de autorregulación	294
4.121 - Ítems del componente «autorregulación»	294
4.122 - Estadísticos de fiabilidad	295
4.123 - Tabla de estadísticos total-elemento	295
4.124 - Medidas de adecuación muestral, ítems 2-28	296
4.125 - Matriz de componentes rotados; ítems 2-28	296
4.126 - Tabla de correlaciones del componente 1, examen de tipo <i>preguntas abiertas</i>	297
4.127 - Tabla de correlaciones del componente 2, examen de tipo <i>test por aciertos</i>	298
4.128 - Correlaciones del componente 1, examen de tipo <i>test basado en la confianza</i>	299
4.129 - Fiabilidad de los agrupamientos «eficacia docente» de cada tipo de examen	300
4.130 - Tabla de correlaciones internas del agrupamiento 3	300
4.131 - Tabla de correlaciones internas del factor 5	301
4.132 - Tabla de correlaciones internas del factor 6	301
4.133 - Estudio de correlaciones internas del componente 7	302
4.134 - Estudio de correlaciones internas del componente 8	303
4.135 - Tabla de correlaciones internas del componente 9	303
4.136 - Correlación entre <i>Nota según aciertos</i> y PC	305
4.137 - IC de cada grupo de estudiantes encontrado en 2008 y 2009	306
4.138 - Estadísticos de la <i>puntuación en confianza</i> y del <i>índice de confianza</i>	307
4.139 - Estadísticos de la <i>puntuación en confianza</i> y del <i>índice de confianza</i>	309
4.140 - Tabla de correlaciones entre las medidas de la autoconfianza y las notas del examen	309, 357
4.141 - IC en varones y mujeres (2009)	311
4.142 - Relación de la confianza con el género de los estudiantes	312
4.143 - Datos estadísticos de la PC, IC y PC-aciertos	312, 362
4.144 - TBC-aciertos, comparación mujeres-hombres	313

TABLAS DEL CAPÍTULO 5	PÁGINA
5.1 - Media de los ítems que definen el constructo « <i>validez auto percibida</i> »	324
5.2 - Correlaciones con los ítems que valoran el tiempo dedicado al estudio con cada examen con la validez autopercebida de cada uno de ellos	325
5.3 - Correlación de la validez de un examen con la necesidad de comprender bien los conceptos estudiados	326
5.4 - «Eficacia docente» de cada tipo de examen (modificada de la 4.58)	331
5.5 - Correlación del factor «eficacia docente» con enfoques de aprendizaje y autorregulación en los tres tipos de examen	332
5.6 - Correlaciones de enfoques profundos con las opiniones sobre el TBC	336
5.7 - Correlaciones de los enfoques superficiales con las opiniones sobre el TBC	337

5.8 - Correlaciones entre la puntuación en enfoques profundos y los dos ítems que miden la <i>habilidad</i> con el TBC	339, 346
5.9 - Correlación con los ítems de enfoques profundos y las puntuaciones del examen	341
5.10 - Correlaciones entre el resultado TBC–aciertos y enfoques, motivos y estrategias profundas	343
5.11 - Correlación entre TBC-aciertos e ítems de estrategia profunda	343
5.12 - Estadísticos de PC–aciertos	356
5.13 - Correlaciones PC–aciertos – notas del examen	359

ÍNDICE DE GRÁFICOS

NÚMERO Y TÍTULO DE LA FIGURA	PÁGINA
FIGURA 1.1. DIAGRAMA DE BLOOM	30
FIGURA 4.1. “PREFERENCIA POR CADA TIPO DE EXAMEN”	188
FIGURA 4.2. “TIEMPO DEDICADO AL ESTUDIO”	196
FIGURA 4.3. “ESFUERZO DE COMPRENSIÓN”	201
FIGURA 4.4. “ESFUERZO DE MEMORIZACIÓN”	206
FIGURA 4.5. “VALIDEZ AUTOPERCIBIDA”	218
FIGURA 4.6. “EFICACIA DOCENTE”	241

CAPÍTULO 1.- INTRODUCCIÓN Y JUSTIFICACIÓN

El autor del presente trabajo ha tenido la fortuna de ser docente de estudiantes universitarios desde hace veinte años. A lo largo de mi vida como profesor me he tenido que enfrentar a todas las cuestiones y problemas que plantea el aprendizaje del alumno, en todas sus dimensiones; uno de los aspectos que más interés me han despertado es la cuestión trascendental de la evaluación del aprendizaje. En estos años de docencia he ido utilizando muy diversos sistemas de evaluación, como preguntas abiertas, trabajos de redacción individuales y grupales, exámenes de respuesta múltiple cerrada (de tipo test), demostraciones orales... Tanto durante mi época de estudiante como en la de docente he conocido el problema apasionante que plantea la evaluación del conocimiento. Algo que sobrepasa, con mucho, la mera necesidad de calificar a los alumnos y darles, o no, licencia para ostentar un título y ejercer una profesión.

La evaluación está en la esencia del proceso de aprendizaje, forma parte de su núcleo más profundo, se inicia en cuanto comienza aquel, se desarrolla a la par que el alumno aprende, culmina cuando alcanza la excelencia, guía las actividades del profesor y está presente en todo momento y ocasión de la relación alumno-docente. Enseñar es evaluar; aprender es autoevaluarse.

Como profesor novato, mis primeras evaluaciones solían ser en forma de preguntas abiertas; pocas y con respuestas largas. Ese procedimiento de examen, el más elemental, me permitía saber mucho sobre los conocimientos del alumno y, también, sobre su proceso de aprendizaje, e incluso, a veces, sobre aspectos de su personalidad o de las características sociales del medio en el que el alumno estaba inmerso. Sigo

pensando que es un buen método de evaluación, aunque tenga defectos y limitaciones, pues es una redacción abierta en la que el alumno tiene plena libertad para expresarse y eso tiene una potencialidad enorme. En la práctica, la gran mayoría de los estudiantes suelen responder a las preguntas de forma estereotipada y memorística, pero algunos escogen una redacción más personal que deja entrever su comprensión de la materia, su forma personal de razonar y hacer suyo el conocimiento, su capacidad de análisis... Cuando eso último sucede, la corrección del examen se convierte en una fuente de satisfacción para el profesor, un auténtico placer que compensa el cansancio y la monotonía general de la tarea.

Pero el procedimiento de evaluación basado en preguntas abiertas de cierta extensión tiene sus limitaciones e inconvenientes, como el autor tuvo pronto ocasión de descubrir. Al gran esfuerzo que supone su corrección se añade el claro fomento de un aprendizaje repetitivo y memorístico. Reproducir exactamente lo que viene en los apuntes o en el libro suele significar un sobresaliente; para conseguir esto, se repite - mentalmente o en alto, según el hábito de cada uno- múltiples veces el párrafo o la lista que hay que aprender hasta que queda fijada, con todas sus palabras y en su orden debido, en la memoria. El que esto escribe fue alumno y sabe de lo que habla. A este defecto general hay que añadir otros, por ejemplo, que, al haber un escaso número de preguntas, el muestreo del conocimiento deje grandes lagunas, lo que se agrava porque, como no todos los temas tienen la misma importancia y el profesor suele hacer mayor hincapié en algunos, gran cantidad de alumnos se los preparan con irregular intensidad, de manera que los temas que más verosímelmente caerán en el examen reciben atención preferente, los otros pueden quedar en blanco; así, el conocimiento que se adquiere tiende a ser parcheado y con numerosas lagunas.

Por todo lo anterior, el autor del presente trabajo decidió utilizar en sus evaluaciones los exámenes de tipo test. Abarcar con amplitud toda la materia, fomentar un aprendizaje menos repetitivo y más centrado en la comprensión y asociación de ideas, esos eran mis objetivos. Y así me introduje en el mundo de las pruebas de respuesta múltiple, donde he ido variando de sistema buscando fijar criterios de evaluación de las respuestas que fueran válidos y fiables.

Uno de los principales inconvenientes que vimos en los exámenes de tipo test era el hecho, que los comentarios de los propios alumnos me revelaban, de que estos exámenes habitualmente motivan un menor esfuerzo de estudio y aprendizaje. De hecho, es el sistema de examen más popular para la mayoría de nuestros estudiantes, y eso no es por casualidad. La idea que flota en el ambiente estudiantil es la de que para aprobar un examen de tipo test se precisa menos tiempo de estudio que para aprobar un examen de preguntas abiertas, ya que las respuestas ya están redactadas y para escoger las correctas existen muchas pistas que ayudan, además del simple conocimiento.

Mi objetivo fue conseguir un modelo de examen que estimulase un mayor trabajo en el alumno y un esfuerzo para analizar, relacionar y comprender mejor los conceptos. En ese sentido, apoyándome en las experiencias publicadas, ensayé varios sistemas de test y, generalmente, escogí poner exámenes combinados de preguntas abiertas y test. La combinación de varios métodos de evaluación, exámenes de tipo test, exámenes de preguntas abiertas, redacciones, trabajos, casos-problema... etc., permite obtener una buena información del nivel de aprendizaje del estudiante desde todos los puntos de vista, pero no siempre el tiempo disponible, el alto número de alumnos y la dinámica del curso permite evaluaciones de esta complejidad. Con frecuencia el profesor escoge el examen de tipo test como el procedimiento más adaptado a la habitual combinación

de alto número de alumnos y poco tiempo disponible para las evaluaciones, aceptando los inconvenientes que hemos mencionado antes. Por todo ello, consideré que sería de interés encontrar un procedimiento de examen de respuestas múltiples que permitiera conseguir la misma calidad evaluatoria que con los complejos exámenes combinados y que fuese suficientemente estimulante del esfuerzo de aprendizaje del alumno.

Cuando me planteé hacer el doctorado en Ciencias de la Educación no me costó mucho decidirme sobre el tema en el que quería trabajar: las técnicas de evaluación. Debo al profesor Pedro Morales la orientación concreta sobre la materia; él fue quien me sugirió que estudiase el sistema de test con puntuación basada en la confianza, él me hizo ver sus potencialidades, a él le debo en buena medida este trabajo, que he ido realizando en los últimos años. Debo decir que mis objetivos de evaluación se cumplieron a mi entera satisfacción con este procedimiento, el test basado en la confianza, pero la demostración de sus ventajas y su gran potencialidad me pareció una materia digna de ser estudiada a fondo. Sistema originario de los Estados Unidos, han sido autores anglosajones quienes más lo han trabajado y estudiado. En el campo de las Ciencias de la Salud, donde desarrollo mi actividad académica, el grupo de Gardner-Medwin en el Reino Unido es quien más ha publicado sobre el test basado en la confianza, y quien ha establecido una mejor escala de calificación de él. Yo he seguido su sistema, aunque con alguna corrección menor. Admitida la viabilidad de este sistema en los países de habla inglesa, yo quería saber hasta qué punto nuestros alumnos, en la unidad docente donde imparto clase, se adaptaban a él. Me interesaba comprobar, en primer lugar, si este procedimiento mantenía la validez y fiabilidad que los autores extranjeros habían comprobado. El estudio de estos puntos lo llevamos a cabo en la parte preparatoria de este trabajo. Comprobada su validez y fiabilidad, he seguido profundizando en el estudio del sistema de evaluación basado en la confianza, que ahora

utilizo de forma sistemática en mis asignaturas. La dirección de mi estudio tenía como objeto el intentar comprobar hasta qué punto este procedimiento de examen modificaba la forma de estudiar del alumno y si la modificación se producía en el sentido que queríamos, el de aumentar el esfuerzo de comprensión. También si este sistema incentivaba los enfoques profundos de aprendizaje o la autorregulación. Considero que sería muy interesante disponer en nuestro sistema educativo de un sistema de examen de tipo test que, siendo bien aceptado y comprendido por los estudiantes, superara las limitaciones de los test tradicionales y demostrara ventajas como instrumento para estimular un mayor y mejor esfuerzo de aprendizaje. Creo que el test basado en la confianza cumple esas características y a estudiarlo he dedicado varios años de trabajo. El resultado de estos años de trabajo es lo que vamos a presentar a continuación.

CAPÍTULO 2.- MARCO CONCEPTUAL Y ESTADO DE LA CUESTIÓN

El estudio de la cuestión lo vamos a dividir en 4 partes:

- En la primera parte expondremos el concepto de *evaluación* y hablaremos del sistema de evaluación de pruebas objetivas de respuesta múltiple, el test; sus características, ventajas y debilidades. Acabaremos exponiendo, brevemente, diversos tipos de test que intentan mejorar las debilidades generales del sistema de examen de tipo test entre ellos el *test basado en la confianza* (TBC) que es el objeto del presente trabajo.
- En la segunda parte expondremos la cuestión de la *confianza académica* y su engranaje con los conceptos de metacognición, autoconcepto, autoeficacia y autorregulación.
- En la tercera parte hablaremos de los *enfoques de aprendizaje*, profundo y superficial, pues su relación con el test basado en la confianza será uno de los campos principales de estudio del presente trabajo.
- En la cuarta parte analizaremos con detalle los diferentes sistemas de examen de tipo *test basados en la confianza* y expondremos las características de la modalidad de TBC que nosotros hemos escogido como objeto de estudio.

2.1. – FINALIDAD DE LA EVALUACIÓN DEL ESTUDIANTE

La característica que mejor define la experiencia vital del ser humano es la de ser toda ella un proceso de aprendizaje. Ese proceso se extiende durante toda nuestra vida, mientras conservamos capacidad de memorización; precisamente la pérdida de la capacidad de aprender señala el inicio del deterioro irreversible de nuestras capacidades intelectuales que llamamos demencia. Dentro de esa larga secuencia vital de aprendizajes, el aprendizaje académico es solamente, por el tiempo que ocupa y por la cantidad de materia asimilada, una parte menor; pero aunque minoritaria es generalmente crucial por su importancia para el individuo. ¿Qué diferencia al aprendizaje académico del resto de aprendizajes que conforman nuestra experiencia vital? Sin duda podríamos decir que la diferencia está en que es un aprendizaje reglado, con procedimientos y objetivos definidos; pero la característica que lo señala y diferencia realmente es la de que es un *aprendizaje evaluado*. La evaluación se constituye así en la característica definitoria más significativa del aprendizaje académico, de la enseñanza reglada y regulada. Como dijimos en la introducción a nuestro trabajo, «enseñar es evaluar, aprender es autoevaluarse»; ello, naturalmente, se refiere principalmente al aprendizaje en la escuela y en la universidad, aunque sin dificultad podríamos extender esa definición, más allá de nuestro entorno docente concreto, al ancho campo de la vida como experiencia global de aprendizaje.

Cuando los profesores estamos elaborando la programación de las materias que impartimos, o cuando, simplemente, estamos haciendo nuestro esquema mental de la organización del curso, solemos colocar en el último lugar de las diversas actividades, al menos desde el punto de vista cronológico, a la evaluación. En la idea habitual que se suele tener, dentro y fuera de la comunidad educativa, la evaluación cierra el

aprendizaje, lo juzga y emite sentencia sobre él. Pero la evaluación no es solamente el resultado final del aprendizaje de una materia o una competencia, si no que forma parte de él desde el principio; hasta el punto de que sin ella, este no tiene pleno sentido. Es imprescindible para los dos sujetos principales del proceso educativo: para el alumno, que necesita conocer sus progresos y deficiencias a fin de llevar a buen término su aprendizaje, y para el profesor, que tiene la responsabilidad de conocer el estado de aprendizaje del alumno y tomar decisiones docentes en función de la información recibida. En otro plano distinto, la evaluación es también necesaria para la sociedad, que requiere garantías sobre la formación de sus profesionales, así como del contenido de los diversos grados y titulaciones. A medida que el aprendizaje se vuelve más complejo, la evaluación se vuelve más complicada, pero adquiere también una mayor importancia (Urosa, 1995). Como dijo el profesor Pedro Morales Vallejo, «lo que no se *evalúa*, se *devalúa*», sin evaluación no hay enseñanza. La manera en que evalúan los profesores dirige la manera en que aprenden los alumnos; por eso, la evaluación no es solamente el remate final del proceso sino que es necesaria a lo largo de todo él, diríamos que casi desde el comienzo. Con un papel tan determinante en el aprendizaje, sorprende el escaso conocimiento y poco interés con el que en ocasiones nos tomamos los que somos profesionales de la enseñanza esta parte trascendental de nuestra tarea.

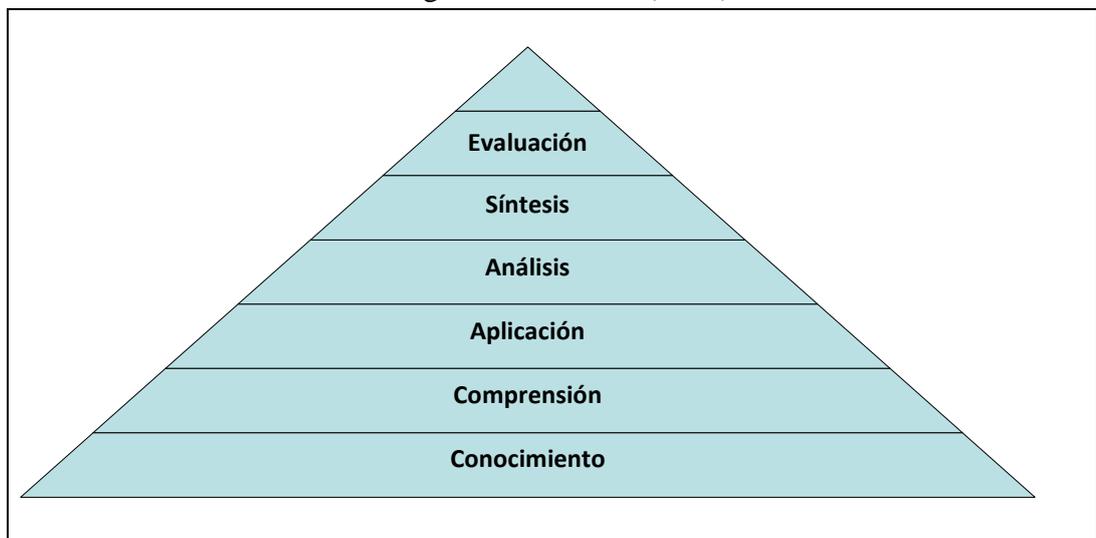
Richard Mayer (2002) define «evaluar» como «hacer juicios basados en unos criterios o estándares previos»; estos criterios el autor los define como *calidad*, *consistencia*, *eficiencia* y *efectividad*. Los estándares pueden, a su vez, ser cuantitativos o cualitativos. Estos juicios pueden ser hechos por el propio estudiante o dados al estudiante por otros.

En función del sujeto que la realiza, distinguimos dos tipos de evaluación, la autoevaluación, del propio estudiante sobre sus conocimientos y aptitudes, y la evaluación que realiza otro, en nuestro caso el profesor, sobre los estudiantes para dirigir y conocer los resultados del aprendizaje. Nosotros pensamos que, aunque ambas son necesarias, la más importante para la adquisición de conocimiento es siempre la auto-evaluación. Para Mayer (2002) la auto-evaluación es la característica propia que mejor define la calidad del aprendizaje; incluye el proceso cognitivo de *comprobar* (referido a los juicios sobre la consistencia interna de lo aprendido) y *criticar* (lo que se refiere a los juicios basados en criterios externos). En este sentido, la tarea del docente en la evaluación es doble: Por un lado, debe hacer la evaluación del aprendizaje del alumno para juzgar si ha sido suficiente, de acuerdo con los requisitos mínimos en conocimientos y competencias que juzga que exige la materia que imparte, y, por otro lado, ha de proporcionar los materiales, pautas y técnicas que ayuden al alumno a autorregular su estudio, a reorientar su aprendizaje y que creen en él una actitud de autoevaluación permanente.

Las evaluaciones no solo estimulan al alumno en su aprendizaje, también le orientan sobre los conocimientos, habilidades y competencias que está adquiriendo y ha de adquirir en el futuro, lo que deducirá de lo que se le exija en las distintas pruebas o exámenes que haya de superar. Es tarea del profesor, además de impartir conocimiento, orientar al alumno sobre qué y cuánto conocimiento ha de adquirir, y los mejores procedimientos para hacerlo. Deberán quedar muy claros los resultados de aprendizaje necesarios, y todo esto solo es posible mediante el proceso evaluativo tutorizado o llevado a cabo directamente por el profesor.

Siguiendo la taxonomía de Bloom (1956), que a pesar de los casi 60 años de antigüedad que tiene no ha perdido validez, la evaluación está en la cima del proceso educativo. Las seis fases que distingue en el proceso son: 1) El nivel de conocimientos del alumno. 2) La comprensión de la información que el alumno posee. 3) La capacidad del alumno para aplicar los conocimientos adquiridos. 4 y 5) La capacidad de razonamiento del alumno sobre la materia, centrada en su capacidad de análisis y síntesis. 6) La evaluación, que en este modelo considerada, fundamentalmente, la capacidad adquirida por el alumno para autoevaluarse. Solo cuando el alumno ha completado todas las fases -desde el mero conocimiento hasta el nivel más alto de dominio, la capacidad de análisis y síntesis- podrá realizar la auto-evaluación.

Figura 2.1.
Diagrama de Bloom (1956)



De manera que para Bloom el nivel más alto del proceso educativo lo constituye la evaluación, por encima de los conocimientos, la comprensión, la aplicación, el análisis y la síntesis. Constituye, en suma, la competencia cognitiva más compleja.

Se han hecho múltiples revisiones de la taxonomía de Bloom, que mantienen su esquema general con pequeñas modificaciones. La más citada de ellas (Anderson y Krathwohl, 2001; Krathwohl, 2002) divide los resultados de aprendizaje en tres niveles: *conocimiento factual*, o de los hechos, *conocimiento conceptual*, de los conceptos y, como nivel superior, *conocimiento procedimental* o de los procedimientos operativos.

Tabla 2.1.

Modificación por Anderson y Krathwohl del diagrama de Bloom

Conocimiento de los hechos	Conocimiento de los conceptos	Conocimiento de los procedimientos			
Recordar	Comprender	Aplicar	Analizar	Evaluar	Crear
Conocimiento metacognitivo					

Nota.(Anderson y Krathwohl, 2001; Krathwohl 2002)

El sentido del progreso en el conocimiento es de la izquierda a la derecha, del mero conocimiento factual hasta el dominio de los procedimientos. La evaluación está en el nivel de *conocimiento de los procedimientos*, en el segundo lugar más avanzado, correspondiendo el puesto superior en la escala a la creación («crear»).

Nosotros entendemos que la evaluación no solo está en el vértice superior del proceso formativo, como en la pirámide de Bloom, sino que aparece en todas sus etapas. La evaluación acompaña al aprendizaje; lo orienta y encamina; corrige sus desviaciones; lo estimula y potencia, y, finalmente, lo juzga. Por medio de ella, el alumno va adquiriendo un dominio mayor de su propio conocimiento. Ella le permite sobrevolar su campo de conocimientos y aptitudes, valorar sus dimensiones, su eficacia, detectar sus lagunas, comprender el sentido general; tener, en definitiva, un *conocimiento de su propio conocimiento*. Esta sería, en nuestra opinión, la más potente definición de la evaluación: es el conocimiento del propio conocimiento -de los

conocimientos y competencias adquiridas en el aprendizaje, expresado con mayor precisión-. En este sentido, la evaluación es elemento fundamental de la *metacognición*.

2.1.1. PROCEDIMIENTOS DE EVALUACIÓN MEDIANTE PRUEBAS OBJETIVAS DE ELECCIÓN MÚLTIPLE

Desde principios del siglo XX se han ido desarrollando los sistemas de evaluación del conocimiento basados en pruebas objetivas de respuestas cerradas de elección múltiple, los exámenes de tipo test. Parece ser que fue Frederick J. Kelly, en 1914, el que tiene el crédito de haber creado el test de elección múltiple (Matheus, 2006). Pronto se utilizó este procedimiento para valorar las capacidades de los reclutas en la I Guerra Mundial, siendo ese suceso histórico el que provocó la aceptación y generalización posterior de este sistema de evaluación.

Los exámenes de tipo test son cada vez más utilizados en todos los ámbitos donde se intentan evaluar conocimientos o capacidades. Lo son por su objetividad aparente, por la posibilidad de abarcar un amplio campo de conocimientos y por la sencillez y rapidez del proceso de corrección. En este tipo de examen el alumno escoge -por ser correcta o incorrecta, según sea el caso- una o varias respuestas prefijadas a una cuestión. En la posterior corrección, se comprueba si la respuesta escogida era la adecuada a la pregunta. Lo sencillo y rápido del proceso, tanto de la realización del examen como de su corrección, permite incluir en el examen una gran cantidad de preguntas, lo que permite abarcar en extensión la materia impartida, con clara ventaja sobre otros procedimientos de evaluación.

Las ventajas de los exámenes de tipo test son evidentes: la objetividad, la posibilidad de hacer un muestreo más amplio de conocimientos y la menor utilización de tiempo en la corrección, especialmente en la corrección por ordenador (Nicol, 2007; Mansell, 2008). Esas ventajas resultan muy importantes en entornos académicos con un número muy elevado de alumnos, como es el nuestro. Nosotros no preconizamos el examen de tipo test como forma exclusiva de evaluación en una situación ideal; más bien lo consideramos complementario de otras técnicas, como preguntas abiertas, casos problema o redacciones, pero en el contexto real de nuestro medio docente, con grupos de alumnos muy numerosos, es la forma más práctica de realizar una evaluación objetiva y amplia de los conocimientos. La posibilidad, debido al elevado número de ítems que admite el procedimiento, de cubrir todo el campo de conocimientos de una asignatura mediante un test es una ventaja incuestionable del sistema. En un examen basado en casos problema o preguntas abiertas solo se podría cubrir la totalidad del campo de conocimientos haciendo un número extravagantemente alto de preguntas; sería un examen demasiado largo, que exigiría un gran esfuerzo al alumno y, posteriormente, una gran cantidad de tiempo al docente para su evaluación.

Aunque el test de elección múltiple ha sido criticado casi desde el tiempo de su estreno, ha tenido, tal vez, su más formidable desafío durante los años 90 del siglo pasado, cuando un número creciente de educadores, guiados por las teorías constructivistas –propugnadas especialmente por Marton y Säljö (1976a,b), Entwistle (1979, 1983) y Biggs (1987, 1993)- defendieron unos métodos de enseñanza y evaluación que enfatizaban las más altas habilidades de razonamiento, frente a la memorización automática de enunciados. Para Paul Pintrich (2002), el test de elección múltiple requiere solo identificar la respuesta correcta, pero no un auténtico dominio cognitivo de la materia, a diferencia de lo que ocurre en un examen de preguntas

abiertas. Todo esto puso en cuestión la validez de los exámenes de tipo test como adecuado sistema de evaluación.

De acuerdo con Gibbs (1999), el aprendizaje adecuado es el que engrana al alumno en la forma de aprender que es exactamente la deseable para la capacidad que queremos desarrollar; así, si queremos cambiar la naturaleza de cómo los alumnos leen, la mejor estrategia podría ser hacer que los alumnos escribieran ensayos: «para escribir un ensayo se necesita leer en torno a un tema para desarrollar un argumento... Otras formas de aprendizaje, como los test de elección múltiple, es muy poco probable que generen ese tipo de lectura» (Gibbs, 1999). ¿Qué actividad de aprendizaje sería aquella en la que un test de elección múltiple fuese la forma deseable?... No siempre es fácil responder a esta pregunta.

Además de esas objeciones teóricas generales, el inconveniente más importante de este tipo de exámenes son las dudas respecto a la validez de sus resultados (Osterlind, 1998). Este de la validez, es un tema que nosotros vamos a estudiar en la modalidad de examen que hemos escogido (*test basado en la confianza*)

Otros inconvenientes que se suelen aducir son:

- La dificultad de redactar enunciados de las preguntas claros y precisos (Rød, 2010).
- La posibilidad de acierto por azar (Downing, 1992).
- La insuficiente discriminación entre los alumnos medianos y los brillantes.
- El no reconocimiento del conocimiento parcial (Rippey, 1973; Morales, 2006a).

- El hecho de que el formato de las preguntas, la redacción de las opciones de respuesta y hasta su orden influyan en la elección de los alumnos (Wolf, 1995).
- La poca destreza que algunos alumnos bien preparados muestran en el momento de enfrentarse a una elección múltiple.
- La opinión extendida de que para este tipo de exámenes es suficiente un conocimiento superficial de la materia (menor esfuerzo de los alumnos).
- La falta de autenticidad del test de elección múltiple, al no ser las tareas del mundo real habitualmente «cuestiones de elección múltiple» (Wiggins, 1990).
- La influencia de la personalidad del alumno al optar entre contestar o abstenerse de hacerlo cuando se le da opción de hacer esto último.

Consejos para una correcta elaboración de los ítems y respuestas en los exámenes de tipo test.

Muchas de estas objeciones no se deben necesariamente al formato de las pruebas de elección múltiple en sí, sino a la baja calidad de las mismas por deficiente redacción de preguntas y respuestas, por escaso número de ítems o por mal diseño general del examen. Así ocurre con la dificultad de redactar enunciados claros y concisos en las preguntas, con la insuficiente discriminación entre los alumnos medianos y brillantes o con la opinión extendida de que solo requieren un conocimiento superficial de la materia. Un adecuado diseño de las preguntas y de las respuestas evitaría esos problemas (Armstrong, 1993; Haladyna, Downing y Rodríguez, 2002). Un examen de tipo test debe ser tan exigente y discriminatorio como cualquier otra técnica de

evaluación. Una buena elección de los niveles de dificultad de las preguntas permitiría la perfecta discriminación de los alumnos más o menos brillantes y evitaría la sensación de que se puede aprobar este examen con un conocimiento superficial. Igualmente, si se saben diseñar bien los ítems, pueden representar con autenticidad las situaciones reales del mundo exterior. Pero todo esto requiere entrenamiento y destreza en la confección de los test. El profesor que opta por este tipo de examen ha de haber adquirido esa destreza necesaria para hacer un buen examen de tipo test, con ítems bien diseñados y claros y un número de preguntas y de respuestas por cada una adecuado; si no es así, es preferible que opte por otros formatos de evaluación. La correcta redacción de las preguntas y respuestas de un examen de tipo test es un proceso exigente, no es sencillo diseñar adecuadamente los ítems; en este sentido, se ha llegado a decir que «la redacción de los ítems [para un examen de tipo test] es una ciencia aún inmadura» (Cronbach, 1970; Haladyna y Downing, 1993). Nosotros echamos de menos una mejor formación del profesorado en esta importante materia. Haladyna y otros (2002), en un trabajo donde hacen una extensa revisión de la bibliografía, establecen 31 reglas para una correcta construcción de los ítems de un examen de tipo test, reglas con las que nosotros estamos globalmente de acuerdo, y a las que intentamos ceñirnos en el diseño de nuestros exámenes. A continuación las exponemos en la Tabla 2.2:

Tabla 2.2.

Taxonomía revisada de las guías para componer los ítems de un test de elección múltiple.

Cuestiones referentes al contenido:
1. Cada ítem debería reflejar un contenido y comportamiento mental específicos, bien explicado en las especificaciones del test.
2. Base cada ítem en un contenido importante a aprender, evitando contenidos triviales.
3. Use material nuevo para evaluar los más altos niveles de aprendizaje. No reproduzca literalmente el lenguaje del libro de texto o el utilizado durante la instrucción para evitar la simple memorización literal.

4. El contenido de cada ítem debe ser independiente del contenido de otros ítems en el test, evitando preguntas encadenadas o con contenidos que se solapen.
5. Evite contenidos demasiado específicos o demasiado generales al redactor los ítems de elección múltiple.
6. Evite ítems basados en opiniones.
7. Evite ítems “tramposos” (preguntas con «truco»)
8. Utilice un lenguaje sencillo, en la medida en que lo permita la material testada, adecuado a los alumnos examinados.
Cuestiones de formato:
9. Utilice las cuestiones, formatos y mejores versiones de respuesta en los formatos de test convencionales, test de elección alternativa, verdadero-falso (VD), múltiple verdadero-falso (MVD), test de emparejamientos, ítems dependientes de contexto y formatos de colocar ítems, pero evitar los formatos de test de respuestas múltiples complejos (tipo K).
10. Utilice un formato vertical para los ítems, mejor que otro horizontal.
Cuestiones de estilo:
11. Revise, compruebe y corrija los ítems
12. Utilice una correcta gramática, puntuación y uso de mayúsculas.
13. Utilice en los ítems textos lo más breves posible..
Tema y contenido de la pregunta:
14. Asegúrese de que el sentido del enunciado es muy claro.
15. Incluya la idea central en el enunciado y no en las respuestas.
16. Evite los enunciados «barrocos» (con excesiva verborrea).
17. Utilice enunciados positivos, evitando los negativos como «no» o «excepto». Si se usan palabras negativas, usarlas con cautela y asegurarse de que la negación queda clara, por ejemplo mediante mayúsculas o poniéndola en negrita.
Elaboración de las alternativas:
18. Puede utilizar tantas respuestas por ítem como se considere necesario, pero las investigaciones sugieren que el número óptimo es el de tres.
19. Asegúrese de que solo una de las respuestas es la correcta (y de que las otras son incorrectas sin ambigüedad)
20. No ponga siempre la respuesta correcta en el mismo lugar, cambiarla entre las distintas elecciones.
21. Ponga las elecciones en un orden lógico o numérico, mejor que aleatorio.
22. Las elecciones deben ser independientes, sin solapamientos.
23. Las elecciones deben ser homogéneas en contenido y estructura gramatical.

24. La extensión de las elecciones debe ser aproximadamente igual.
25. « <i>Ninguno de los anteriores</i> » debe ser usado con precaución.
26. Evite la respuesta « <i>Todos los anteriores</i> ».
27. Las frases deben ser positivas, evitar negaciones como «NO»
28. Evite dar claves sobre la respuesta correcta, como: <ul style="list-style-type: none"> a. Determinantes específicos como <i>siempre, nunca, completamente</i> o <i>absolutamente</i>. b. Asociaciones lingüísticas, como elecciones idénticas o con palabras muy similares al enunciado. c. Inconsistencias gramaticales que puedan dar la clave de la respuesta correcta. d. Respuestas correctas demasiado conspicuas. e. Pares o tripletes de opciones que den claves sobre la respuesta correcta. f. Opciones absurdas o ridículas.
29. Elabore distractores verosímiles.
30. Use errores típicos de los estudiantes para elaborar los distractores.
31. Utilice el humor con precaución, solo si es compatible con el entorno docente y el profesor.

Nota. Haladyna y otros (2002)

Dada la importancia de este punto, la calidad en la elaboración de los exámenes, creemos que merece la pena detenerse en estas reglas para dar nuestra opinión, fundada en la experiencia que hemos acumulado en la docencia.

Nosotros secundamos totalmente la mayoría de las recomendaciones de Haladyna, Downing y Rodríguez. Por ejemplo, en cuanto al contenido, nosotros desaconsejamos totalmente las preguntas encadenadas o con solapamiento de contenidos, que hacen que el estudiante que falle la primera de ellas tienda a fallar las sucesivas, exagerando la penalización por un defecto cognitivo concreto (punto 4). Nos parece fundamental, en nuestras materias de Ciencias de la Salud, que intentamos que sean basadas en la evidencia, evitar los ítems basados en opiniones (punto 6).

Uno de los errores que hay que evitar siempre es el expuesto en el punto 7, evitar preguntas tramposas o con truco; este tipo de preguntas, que desgraciadamente

seguimos viendo en algunos exámenes, tienen un doble efecto pernicioso: por un lado inducen muchos fallos no debidos a falta de conocimiento sino al exceso de confianza del alumno; por otro lado, cuando el alumno identifica que «hay truco» en algunas preguntas tiende a desconfiar de los enunciados de todas las demás, buscando significados no literales –«trucos»– en ellas, incurriendo en los consiguientes errores de interpretación de los enunciados y fallos forzados.

El punto 8 nos dice que hemos de usar un «vocabulario simple»; creemos que se refiere a que el vocabulario ha de estar en línea con el que se ha utilizado durante las clases expositivas y en los libros de texto recomendados; evitando cultismos innecesarios.

La elaboración de enunciados adecuados en las preguntas es una tarea nada fácil que exige del profesor una importante cualidad, le exige *empatía* con el alumno. El examinador ha de saber pensar como suele pensar el estudiante, verbalizar los conceptos tal como él los verbaliza, y, en definitiva, ponerse en su lugar al enfrentarse a las preguntas del examen. Muchas veces nos ha ocurrido que, al reclamar los resultados de algunas preguntas falladas en el examen, el estudiante nos enseñaba a interpretar los enunciados desde un punto de vista diferente al nuestro, interpretación alternativa pero también correcta, que nos obligaba a anular la pregunta en cuestión por tener un significado ambiguo. No es fácil ponerse en el lugar del estudiante, pensar como él lo suele hacer, pero es necesario si se quiere hacer un test válido.

Entre los puntos correspondientes al formato y el estilo, es importante el punto 11: revisar y comprobar exhaustivamente los ítems. Elaborar un examen de tipo test es una tarea compleja que exige tiempo y varias revisiones. En este sentido, es evidente que se

ha de ser muy cuidadoso en no cometer nunca faltas gramaticales, de puntuación o incorrecciones sintácticas, lo que daría una impresión penosa y sería *deformativo* para el estudiante (punto 12); una revisión cuidadosa gramatical es necesaria antes de entregar el examen a la imprenta. Un consejo importante, pero que muchos examinadores no siguen, es el de reducir al máximo la extensión de los enunciados de las preguntas, disminuyendo el esfuerzo y tiempo que el examinando tiene que dedicar a su lectura e interpretación (punto 13); nosotros también aconsejamos enunciados breves y fáciles de interpretar, para disminuir la posibilidad de errores no debidos a falta de conocimiento.

El sentido de la pregunta que se hace debe ser muy claro, sin ambigüedades o conceptos borrosos (punto 14). El punto 15 hace referencia a que la idea central que se cuestiona debe figurar en el enunciado y no en las respuestas; esto hace referencia a ítems de este tipo: «¿Cuál de las siguientes frases es la correcta?...»; si bien alguna pregunta así es admisible –estando bien diseñada– nosotros también creemos que en un test bien construido el enunciado de la pregunta debe contener la idea central y las respuestas deben ser eso: simplemente respuestas. En el punto 16 se indica la necesidad de evitar enunciados con excesiva verborrea («barrocos»).

Importancia especial tiene el punto 17; los autores (Haladyna y otros) propugnan evitar abusar de los enunciados negativos (por ejemplo: «¿Cuál de las siguientes respuestas no es correcta?...» «todas menos...»); si se decide utilizarlos, es conveniente destacar el carácter negativo de los enunciados usando en la negación mayúsculas o negritas. En este punto nosotros coincidimos totalmente con las recomendaciones de los autores. En nuestra experiencia los enunciados negativos inducen equivocaciones no debidas a la ignorancia en los estudiantes; sin conocer las recomendaciones de Haladyna y otros, nosotros espontáneamente ya habíamos reducido al mínimo estas

cuestiones y resaltado la negación aumentando el tamaño de la letra o con negrita, y aún así detectábamos algunos errores (alumnos que fallaban estas preguntas y cuyo dominio de la materia, puesto de manifiesto en las demás preguntas del examen, hacía ilógicos estos fallos). Por otro lado, el sistema de examen que utilizamos desde hace años, y que es analizado en el presente trabajo, no es nada adecuado para este tipo de preguntas, que aumentan la incertidumbre y reducen la seguridad del examinando. Por ello, no incluimos nunca más de un 5 % de estas preguntas de enunciados negativos (1, 2, o a lo sumo 3, entre 60).

Los consejos referentes a las respuestas son compartidos por nosotros en su totalidad. Así, en el punto 18, escogemos solo tres opciones de respuesta. Aunque hay algunas preguntas en las que fácilmente se pueden encontrar tres o más alternativas falsas («distractores») verosímiles, si se escogen cuatro o más alternativas de respuesta –una correcta y tres o más falsas- suele ocurrir que no es fácil encontrar suficientes distractores plausibles, y los inverosímiles son fácilmente detectados por el alumno como *falsos*, aunque este tenga muy pocos conocimientos de la cuestión; son respuestas «de relleno», y su inclusión rebaja la dificultad del examen. Tres respuestas, una correcta y dos falsas, o un test verdadero/falso son las opciones que los autores y nosotros, en nuestra experiencia, recomendamos. Los puntos 19 y 20 son de lógica elemental; en cuanto al 19 hay que dar seguridad al examinando de que solo hay una respuesta correcta –la pregunta «¿puede haber más de una respuesta correcta en esta pregunta?» es hecha sistemáticamente por algún alumno, o varios, en cada examen-. Igualmente hay que variar la posición de las respuesta correcta dentro de las elecciones en las sucesivas preguntas (punto 20); nosotros añadiríamos que hay que variar la posición pero **nunca sistemáticamente**, sino al azar. Es importante que el alumno no

llegue a detectar pautas que le lleven a pensar que la posición de la respuesta correcta en una pregunta condiciona la posición en la siguiente.

Un factor que puede dar pistas al alumno, pistas correctas o falsas –en ambos casos igualmente perjudiciales-, es el orden de colocación de las respuestas, cuando no sigue un patrón lógico. Por ello, los autores preconizan que, siempre que sea posible, las respuestas se coloquen en un orden lógico o, cuando sean cuantitativas, numérico (creciente o decreciente); así los examinandos no pensarán que el orden esconde alguna pista (punto 21).

El consejo número 22 es muy relevante: las elecciones que se ofrecen en cada pregunta deben ser independientes y no solaparse; el solapamiento puede conllevar que haya dos respuestas que son, al menos parcialmente, correctas, creando confusión.

La homogeneidad de las respuestas en extensión, gramática y contenido es necesaria para evitar dar pistas –fundadas o no- sobre la elección correcta (consejos 23 y 24). A veces es muy difícil elaborar respuestas de la misma extensión, y para hacerlo puede ser necesario incluir palabras superfluas en una respuesta, lo que no es una buena decisión; en estos casos nosotros preconizamos que todas las respuestas –tres, cuatro o las que sean- tengan diferente extensión, y que sean muy homogéneas en el lenguaje y contenido, para que el alumno no se sienta inclinado a pensar que la diferencia de extensión esconde algo.

Los consejos 25 y 26 hacen referencia a la inclusión de opciones de respuesta como «ninguna de las respuestas anteriores es correcta» o «todas las anteriores correctas». Los autores las desaconsejan, relativamente en el caso de «ninguna...» (recomiendan que se use con cuidado) y absolutamente en el caso de «todas...». La

experiencia nuestra va en el mismo sentido y, así, ponemos pocas veces la opción «ninguna es correcta» y nunca la de «todas son correctas». Ambas respuestas transforman la pregunta de respuesta múltiple en un formato de opciones «verdadero/falso», cambiando el sentido del test. Si se incluye la última opción -«todas correctas»- habiendo más de tres respuestas al alumno se le dan pistas suplementarias, pues si detecta dos respuestas verdaderas sabe, aunque ignore las demás, que la elección ha de ser necesariamente «todas son correctas». Así como los autores recomiendan evitar los enunciados negativos, también recomiendan evitar las respuestas negativas (punto 27).

El punto 28 incluye una batería de consejos para evitar dar pistas sobre cuál es la opción correcta. El alumno que no confíe totalmente en sus conocimientos buscará afanosamente pistas de todo tipo en las respuestas; hay que evitar dárselas e, incluso, evitar que parezca que hay pistas, aunque realmente no las haya. Toda pista, sea auténtica o falsa, distorsionará el resultado del examen. Entre los seis consejos que se incluyen en este punto, hay que destacar el de evitar pares –o tripletes- de respuestas; por ejemplo, dos respuestas mutuamente excluyentes y que entre ambas cubran todo el campo de posibilidades, haciendo a las demás superfluas. El último consejo de este punto hace referencia a algo que nosotros consideramos un vicio en la elaboración de los exámenes de tipo test: incluir alguna respuesta obviamente imposible, absurda y, sobre todo ridícula; esto se suele hacer porque no se encuentran tantos distractores plausibles como opciones de respuestas falsas, de ahí la conveniencia de poner solo tres respuestas por pregunta. Las respuestas absurdas transforman una pregunta con N respuestas en otra con N-1 respuestas, pero puntuada como si tuviera todas las respuestas verosímiles. Y dentro de las respuestas poco verosímiles, es necesario evitar

sobre todo las respuestas ridículas; los alumnos con frecuencia las interpretan como una falta de respeto.

El consejo 30 -«utilizar los errores típicos de los estudiantes como distractores»- es muy útil, en nuestra opinión. La experiencia de exámenes anteriores le indica al profesor cuáles son los errores en los que el alumno incurre con más frecuencia; ese conocimiento nos indica los puntos débiles del aprendizaje, nos resulta útil para mejorar la docencia, y la utilización de esos errores para elaborar distractores, refuerza esos mencionados puntos débiles. No podemos dejar de recalcar que las evaluaciones no solo sirven para calificar a los estudiantes, son una fuente de información muy valiosa para mejorar el aprendizaje del alumno y la capacidad docente del profesor.

Finalmente, el último consejo, referente al uso del humor en los exámenes, indica que debe ser utilizado con precaución y cuando sea compatible con el ambiente de aprendizaje. Nosotros creemos que se debe ser muy cauto en el uso del humor en el contexto del examen. Nos parece excelente el humor en los demás momentos del proceso educativo, crea un ambiente académico distendido y, bien usado, propicio a despertar el gusto del alumno por el aprendizaje. Pero el momento de un examen no es muy propicio para el humor. Simplemente, los alumnos «no están para muchas bromas». Si bien una moderada utilización del humor, cuando sea pertinente, puede relajar el ambiente y disminuir la tensión –eso sería beneficioso en un examen-, creemos que debe limitarse al momento de las instrucciones y consejos generales del examen y nunca figurar en el texto o cuerpo del examen. Algunos alumnos lo podrían agradecer, pero nuestra experiencia nos indica que siempre hay alumnos que lo consideran una falta de respeto, una burla fuera de lugar y ocasión.

Estos son los consejos dados por Haladyna y otros en su comunicación publicada en 2002. Aunque se puede discrepar de algunos puntos, cada uno tiene su propio estilo de elaborar exámenes, los consideramos muy adecuados para mejorar la calidad de las evaluaciones. Lo importante es destacar que la elaboración de un examen de tipo test es un proceso muy exigente, que pone a prueba la capacidad del profesor como docente. Hay que dedicar tiempo a formarse en las técnicas de evaluación, hay que dedicar tiempo y esfuerzo a cada examen y nunca ser rutinario. El examen es un arma de importancia crucial para el aprendizaje, debemos saber utilizarla bien.

En nuestro caso, en que utilizamos el test basado en la confianza, hemos de llevar al extremo el cuidado en la elaboración de cada examen. Especialmente hay que esforzarse en que la elección de ítems y las respuestas no creen incertidumbres en el estudiante, pues las incertidumbres y dudas condicionan, en nuestro sistema, la nota final. Nosotros usamos preguntas positivas, con solo tres respuestas. Intentamos que las respuestas sea sin solapamientos, homogéneas y sencillas al máximo en su lectura e interpretación, sin opciones de «todas correctas», sin respuestas negativas, sin distractores absurdos, sin trampas y con todas las respuestas igualmente plausibles. Procuramos que no haya nunca, por acción u omisión, alguna pista. Vigilamos el lenguaje, intentando evitar inconsistencias y ambigüedades en la interpretación. Y, a pesar de todo esto, en la mayoría de nuestros exámenes hemos de anular alguna pregunta al descubrir con posterioridad que era defectuosa en su enunciado o en sus respuestas.

2.1.1.1. OTRAS CRÍTICAS A LOS EXÁMENES DE TIPO TEST

Otro motivo de crítica a este tipo de exámenes es el hecho de que pueda estimular en el alumno un aprendizaje superficial (Gardner-Medwin, 2006). Como dijimos antes, la pregunta correcta recibe la misma puntuación, la máxima para cada ítem, domine el alumno la materia, tenga dudas sobre ella o simplemente la acierte por adivinación. Esto no ocurriría en otros modelos de evaluación distintos del examen de tipo test; por ejemplo, en un examen de preguntas abiertas, el alumno expresará su grado de dominio o su falta de él en la materia al redactar la respuesta, y si carece de conocimientos sobre el tema preguntado no existe posibilidad de «acierto por azar». En definitiva, en un test, el examinando con frecuencia obtendrá la misma puntuación tenga un conocimiento exhaustivo de la materia o solo una idea superficial y aproximada, ya que las respuestas están redactadas y los factores que le llevan a escoger la correcta no siempre son profundos. Damos por hecho que el conocimiento superficial de una materia inducirá en el alumno más respuestas incorrectas que el conocimiento profundo, pero no tenemos garantías de que esto suceda realmente así y, sobre todo, de que esto suceda en la proporción correcta. Por ello, los exámenes de tipo test pueden incitar al alumno a desarrollar un aprendizaje superficial (Gardner-Medwin, 2006). Para contrarrestar esto, el profesor tiene siempre la capacidad de poner en el examen preguntas que exijan un alto dominio de una materia, pero las preguntas difíciles con frecuencia tienen enunciados complejos que multiplican las posibilidades de confusión en el alumno y de error no debido a la falta de conocimiento; no es este un tema fácil de resolver. En cualquier caso, en un examen de tipo test bien diseñado siempre debe haber una combinación de preguntas de diferentes grados de dificultad.

Al adquirir progresivamente más experiencia de exámenes, los estudiantes pronto aprenden, al menos con las preguntas menos difíciles, que pueden acertar confiando en asociaciones superficiales de conceptos, lo que conlleva un escaso incentivo para pensar rigurosamente o desarrollar un conocimiento profundo; con frecuencia el resultado es el mismo: *habitualmente la primera respuesta que ellos piensan es la correcta* (Gardner-Medwin, 2006). En definitiva, la experiencia docente en un ambiente universitario indica que los alumnos se plantean de forma diferente la preparación de los exámenes cuando son de tipo preguntas abiertas o casos problema que cuando son de tipo test; en este último caso el estudio suele ser más superficial y ocupar menos tiempo.

Entre los reproches habituales referidos a las pruebas objetivas cerradas de elección múltiple hay dos en los que se suele hacer el mayor hincapié y, por ello, nos interesa especialmente destacarlos: la **no valoración del conocimiento parcial** y la posibilidad de **acierto por azar**.

2.1.1.2. NO VALORACIÓN DEL CONOCIMIENTO PARCIAL EN LOS TEST

Todas las variedades de exámenes de tipo test se reducen, en última instancia, a una serie de enunciados a los que hay que calificar como «*verdaderos*» o «*falsos*». Este esquema es invariable, sea el modelo de test con solo dos respuestas, explícitamente *verdadero/falso*, o de respuestas múltiples, en donde el alumno ha de decidir en cada una de las opciones si es correcta o incorrecta. Es válido tanto para esquemas de exámenes con varias opciones correctas en cada ítem (Pomplun y Omar, 1997), como para los más tradicionales, en los que hay una sola respuesta correcta. Esta dualidad característica de estas pruebas *-respuesta correcta/incorrecta-* excluye las matizaciones y no prevé grados intermedios de conocimiento (Morales, 2006a).

Los test convencionales de elección múltiple son un caso muy especial dentro de una situación mucho más general de toma de decisiones. Al sujeto se le proporciona un cuerpo de información, un conocimiento del campo, un número limitado de opciones para escoger, y este desarrolla un algoritmo de preferencias haciendo una elección simple entre las opciones de que dispone (Rippey, 1973). Esa decisión, sin embargo, no es muy instructiva sobre el estado de conocimiento del individuo. La elección única, sin matices, no separa al sujeto confiado del dubitativo. Nada distingue al jugador afortunado del experto cualificado y cierto, salvo nuestra suposición previa de que el segundo acertará más preguntas que el primero; pero ante una pregunta respondida correctamente, no podemos saber si el estudiante la ha acertado por tener un conocimiento cierto y fundamentado, un conocimiento parcial o por haber tenido suerte. Al haber una única respuesta correcta se elimina el área extensa entre el conocimiento garantizado y la respuesta aleatoria (Rippey, 1973). Las matizaciones pueden figurar en el enunciado de las preguntas, pero las respuestas son duales y excluyentes; se escoge o no se escoge cada una de las alternativas de respuesta.

En la realidad, sin embargo, se dan con frecuencia situaciones en las que nuestro grado de conocimiento no nos permite una total seguridad a la hora de escoger la alternativa, un 100 % de probabilidad de acertar, sino solo una probabilidad menor; y podríamos hacer aseveraciones de este estilo: «*muy probablemente correcto*», «*probablemente...*», «*poco probable...*». En esas condiciones, habituales en los exámenes y, desde luego, también en las elecciones de nuestra vida no académica, el sujeto ha de asumir un riesgo, ha de hacer una elección con posibilidades de no acertar; en suma, ha de fiar parte de su acierto o puntuación al albur de un conocimiento incompleto de la cuestión.

Los exámenes convencionales de tipo test, que puntúan según el número de aciertos y fallos, no suelen contemplar adecuadamente en las formas más utilizadas las situaciones intermedias -de conocimiento incompleto, de posibilidad de acertar o fallar por la suma de conocimientos y azar- (Morales, 2006a); en ese sentido se separan de la vida real, en la que es frecuente, probablemente *lo más frecuente*, que se dé un conocimiento incompleto, parcial o simplemente inseguro de las múltiples variables que determinarán el acierto de nuestras elecciones.

En un examen de tipo test de elección múltiple, el conocimiento parcial es la situación que se da cuando, al buscar la opción correcta, el estudiante no es capaz de encontrarla positivamente, pero puede descartar una o varias opciones por erróneas. Si su conocimiento parcial es atinado, ello le permite aumentar sus posibilidades de acertar la pregunta, aunque ignore la respuesta correcta. Así, por poner un ejemplo, en un test con tres opciones de respuesta, si el alumno desconoce cuál es la correcta pero sabe que una de ellas es definitivamente incorrecta, sus posibilidades de acertar por azar, descontando esa respuesta, se elevan de un 33 % a un 50%; le resultaría rentable, en términos de puntuación del examen, responder a esa pregunta. En los sistemas de test convencionales, queda a la elección del alumno el responder o no; tomará su decisión en función de factores que nada tienen que ver con su competencia en la materia examinada; responderá a la pregunta o se abstendrá según su personalidad –mayor o menor tendencia a sumir riesgos-, según su situación docente personal, su cálculo estratégico, sus aspiraciones de puntuación, etc. No nos parece adecuado que estos factores personales y psicológicos, factores espúreos, influyan en la puntuación de una asignatura; además, pueden influir de una forma imposible de controlar en cada alumno. Nosotros propugnamos, para evitar esa influencia de factores extra-cognitivos, que en los exámenes convencionales no se le dé al alumno la posibilidad de abstenerse al

contestar a una pregunta. Como éste puede, en cualquier caso, dejar preguntas sin responder, la forma de conseguir que responda a todas es no penalizar por pregunta fallada, exigiendo para aprobar un número de aciertos que compense el efecto de la adivinación. Como veremos más adelante, hay varios sistemas de examen, entre ellos el que nosotros hemos utilizado –test basado en la confianza- que abordan de forma racional la evaluación del conocimiento parcial.

De la misma manera que el conocimiento *cierto*, que el alumno cree tener, puede ser atinado o equivocado –«conceptos erróneos»-, el conocimiento parcial también puede ser equivocado. Por ello, el estudiante que se enfrenta a una pregunta de examen se encuentra en una de las siguientes cinco posibilidades (Lau, Lau, Hong y Usop, 2011): Conocimiento pleno correcto, conocimiento parcial correcto, ausencia de conocimiento, conocimiento parcial equivocado y conocimiento completamente equivocado. Dadas las múltiples situaciones diferentes en que el alumno se encuentra al afrontar una pregunta de test, el sistema dicotómico de puntuación es insatisfactorio (Hutchinson, 1982).

2.1.1.3. POSIBILIDAD DE ACIERTO POR AZAR

En las pruebas objetivas de respuestas múltiples, el alumno puede escoger la opción correcta por tener un conocimiento cierto o parcial de la materia, o directamente por azar, cuando ignora completamente cuál es la respuesta correcta (Haladyna, 2004). Por añadidura, como hemos señalado antes, la adivinación puede venir de una defectuosa construcción de los ítems (Cronbach, 1970; Haladyna y otros, 2002; Rød, 2010), con enunciados poco precisos o demasiado complejos, o porque contengan

claves que ayuden al alumno, como distractores mal redactados u obviamente imposibles,.

En general, las críticas más numerosas a los exámenes de tipo test provienen de esta posibilidad que tiene el alumno de acierto por azar, por la correspondiente influencia de la suerte (Kurz, 1999). Aparte del efecto distorsionante sobre la evaluación del alumno, algún autor (Oh, 2004) expresa que la adivinación, en parte por azar y en parte por otros factores que no sean el conocimiento cierto, como la defectuosa construcción de los ítems con pistas para el alumno, es una práctica empobrecedora del proceso educativo que interfiere con la auténtica habilidad del estudiante para responder a un test. El alumno que acierta una pregunta tiende a creer posteriormente que si eligió la respuesta correcta era porque se la sabía, aunque en la realidad, en un proceso lleno de incertidumbres, haya escogido la respuesta que le parecía simplemente más verosímil, cifrando una parte importante de la posibilidad de acierto a la adivinación. Esta práctica favorece hábitos descuidados de aprendizaje y es considerada antipedagógica (Oh, 2004; Gardner-Medwin, 2007).

Pese a la importancia que muchos docentes le otorgan, la tan mencionada posibilidad de acierto por azar, se minimiza, y llega a desaparecer en la práctica, con un número suficientemente elevado de preguntas (Downing, 1992). Un número mínimo de preguntas en el test parece necesario para reducir el efecto de la adivinación por azar y aumentar la fiabilidad. Nosotros creemos que se puede conseguir una adecuada fiabilidad y minimizar el efecto del azar con un número suficiente de preguntas (utilizamos de 40 a 60 preguntas en los test de elección múltiple con 3 respuestas). Además del número suficientemente elevado de ítems, una práctica que puede estimular al alumno a abandonar la adivinación en beneficio de un mayor esfuerzo de

razonamiento sería la de premiar el acierto por conocimiento parcial frente al acierto por azar (Bush, 2001, p. 157); no otra cosa es lo que se pretende con el método del test basado en la confianza (Gardner-Medwin, 2006 a).

En el examen de preguntas abiertas a desarrollar, el alumno ha de redactar la contestación; no puede expresar aquello que ignora; la forma de expresión y exactitud en los conceptos expuestos dan al examinador una idea precisa sobre su nivel de dominio de la materia, más allá de que se dé la respuesta correcta a la pregunta. Si el examen es oral, el tono de la voz y la seguridad o dudas del alumno al responder, incluso la expresión corporal, proporcionan pistas adicionales sobre su nivel de aprendizaje y dominio de la materia. Pero en el examen de tipo test, el alumno que acierta la pregunta con pleno conocimiento de la materia y el que la acierta por azar, reciben la misma puntuación. Este es el problema que más suele preocupar a los examinadores que se plantean optar por este tipo de examen: el conocimiento cierto, el conocimiento inseguro o el azar, reciben la misma puntuación si el alumno acierta la pregunta (Hutchinson, 1982). Un elemento que es responsable de esta crítica es el hecho de que, en una prueba objetiva de elección múltiple, el alumno tiene confeccionadas ya las respuestas y no tiene que redactarlas, por lo que puede acertar aunque ignore totalmente ese campo de conocimiento y sea incapaz de redactar por sus medios una respuesta.

Nosotros no compartimos esos temores al efecto de la adivinación por azar; al menos no los compartimos si el examen de tipo test está bien diseñado. En primer lugar, el azar y el juego de adivinación existen en todo examen o sistema de evaluación. Cuando el alumno se prepara una evaluación basada en preguntas abiertas, en casos problema o en una demostración de habilidades prácticas -por citar tipos de evaluación

a los que no se les suele achacar el efecto del azar-, en realidad juega con la adivinación, a no ser que su preparación sea tan excelente que conozca o domine completamente todo el campo cognitivo del examen. Intentará prever qué preguntas, casos o pruebas concretas pondrá el profesor en el examen; lo hará basándose en la experiencia de otros años, en los consejos de compañeros o en sus impresiones durante las clases, y dedicará un esfuerzo preferente a esos campos *privilegiados* en detrimento de aquellos que crea menos proclives a ser evaluados («esto cae seguro», «este tema no va a caer en el examen»...). Por lo tanto, todos los alumnos que no tienen un nivel de preparación perfecto suelen jugar con la adivinación. Las pruebas objetivas de elección múltiple suponen una ventaja a este respecto, pues como los ítems son numerosos, pueden abarcar todo el campo de conocimientos sin dejar lagunas significativas. El alumno, aun así, puede prever qué temas serán los que reciban más atención del profesor, y estén presentes en mayor número de preguntas, pero ya no podrá esperar que haya campos del conocimiento que no se evalúen en el examen, por lo que tendrá un estímulo para que su estudio sea más sistemático y abarque toda la extensión de la materia.

Como dijimos antes, entendemos que el tema del acierto por azar es algo que se controla con el número de ítems del examen. Simplemente escogiendo el número de ítems, hacemos que la influencia del azar en la nota del examen se reduzca hasta los niveles que deseamos. En ese sentido, Downing (1992) propugnó que en el formato de test con dos opciones de respuesta (verdadero/falso) sería adecuado que hubiera, al menos, 100 preguntas; con esta cifra de 100 preguntas en un test con formato verdadero/falso, el autor calculó que la posibilidad de acertar un 70 % por azar sería de solamente de un 0.0035 –o de un 0.35%- (Downing, 1992). La regla que propugna Downing para escoger el número óptimo de preguntas en un test sería aquella que consiste en dividir 200 entre el número de respuestas de cada ítem; la aplicación de esta

regla equivaldría a unas 66 preguntas en un test con tres respuestas por ítem, o de 50 preguntas si las respuestas son 4. Estas cifras serían las ideales si se quiere que el efecto del azar altere la nota menos de un 5 % en cualquier caso, pero pueden resultar algo elevadas y se pueden reducir en la práctica. Nosotros, basándonos en nuestra experiencia, daríamos las siguientes como cifras mínimas de preguntas en un examen de tipo test: 60 ítems para preguntas verdadero/falso, 40 ítems para preguntas con tres opciones de respuesta, 30 para preguntas con cuatro respuestas; por debajo de esas cifras, el efecto del azar en la nota del examen podría ser superior a un 10 %, con un margen de error de un 10 %, aproximadamente, lo cual no lo consideramos aceptable. Esa sería, en nuestra opinión, la cantidad de preguntas mínimas que según el número de respuestas debería tener un examen de tipo test para minimizar el acierto por azar; naturalmente, es preferible aumentar el número de preguntas sobre ese mínimo, sin llegar a poner exámenes que por su elevado número de ítems resulten agotadores para el alumno y propicien errores de este. Nosotros, en nuestra práctica habitual, preferimos utilizar modelos de test con 3 respuestas por ítem, que parecen preferibles a aquellos que tienen 4, 5 o más respuestas (Haladyna y otros, 1993; Rodríguez, 2005), y para estos exámenes con tres opciones solemos utilizar una cifra de preguntas que oscila entre 45 y 60. Luego, a la hora de establecer la puntuación, hay que descontar del total de preguntas y de aciertos el número más probable de aciertos por azar, el 50 % en preguntas verdadero/falso, el 33 % si hay tres respuestas, el 25 % con cuatro, etc.

No obstante lo anteriormente dicho, a la mayoría de los profesores les preocupa el efecto de la adivinación y el acierto por azar. Para evitar este problema, el del acierto por azar, suelen utilizarse distintos procedimientos, que exponemos a continuación.

2.1.2. SISTEMAS DE CORRECCIÓN DE LA ADIVINACIÓN EN LOS EXÁMENES DE TIPO TEST

El procedimiento más sencillo de calificación de un examen de tipo test consiste en lo que llamamos «puntuación según número de aciertos»; en él se puntúa según el número de preguntas acertadas, o, más bien, según el porcentaje de estas sobre el total. Ese fue el sistema de puntuación en los primeros test utilizados, hace ya un siglo. Pronto se tuvo en cuenta la cuestión de la posibilidad de acierto por azar, que tendía a hacer que la puntuación fuese siempre más elevada que la real.

Para corregir el efecto de la adivinación por azar se idearon procedimientos diversos. El más utilizado es el de la llamada *fórmula clásica de corrección de la adivinación*, que consiste en descontar de las respuestas correctas una penalización por las supuestamente acertadas por azar, cuyo número se deduce a partir de las respuestas incorrectas mediante una sencilla fórmula. La penalización de las respuestas incorrectas busca, por un lado, neutralizar el efecto del acierto por azar, y por otro lado, pretende que el alumno no se arriesgue a responder si no conoce la respuesta. Es decir, busca calificar al alumno según la estimación del número de aciertos que no se deben al azar, y también busca evitar que el alumno responda al azar al desincentivar la respuesta si el alumno no está seguro. Fácilmente se puede ver que estos dos objetivos simultáneos encierran una posible contradicción, como explicaremos luego. La fórmula clásica de corrección para eliminar la adivinación es: $\text{respuestas correctas} - \text{respuestas incorrectas}/(n-1)$, siendo “n” el número de opciones de respuesta por cada pregunta del test. Una sencilla demostración matemática, que incluimos en el anexo A, prueba que esta fórmula compensa (sería más correcto decir que *tiende a compensar*) el aumento de puntuación por aciertos al azar; pero **esto solo sucede así si el alumno contesta a todas las preguntas** (ver anexo A).

En este proceso de penalizar las respuestas incorrectas hay un problema intrínseco que la mayoría de los docentes no se plantean, y que es trascendente; es el de qué fin último se busca con la penalización por respuesta incorrecta. ¿Se pretende disuadir al alumno de practicar la adivinación y forzarle a que responda solo a las preguntas que tiene seguras? ¿O se pretende corregir matemáticamente el efecto de la adivinación? Se puede argüir que se persiguen ambos fines a la vez, probablemente eso piensen la mayoría de los que aplican este sistema, pero la cuestión es que ambos objetivos, a la vez, **son incompatibles**, como vamos a tener ocasión de demostrar. Como hemos dicho antes, la fórmula clásica de corrección de la adivinación compensa los aciertos por azar solamente si se contesta a todos los ítems; si el alumno escoge la opción de dejar preguntas sin contestar, entonces esta fórmula «*sobrepenaliza*» la nota. La demostración matemática de esta última afirmación, la «sobrepenalización» cuando se utiliza la fórmula clásica de corrección de la adivinación, es muy sencilla (la hemos incluido en el anexo A). Básicamente, el sistema considera que toda pregunta dejada en blanco se debe a desconocimiento completo de la materia; pero es posible que el alumno la deje en blanco teniendo un conocimiento parcial, por no querer asumir riesgos; en ese caso, renuncia a unas probabilidades de acertar que compensan en exceso la penalización. Y si el alumno desconoce completamente la pregunta y se arriesga a responder al azar, el sistema prevé una penalización que anula exactamente el azar, y, por lo tanto, tiende a una penalización nula por la adivinación pura. Por lo tanto, si un examen incluye este tipo de penalización por respuesta incorrecta, el alumno inteligente debería contestar a todos los ítems, supiera o no supiera la respuesta, pues así podría obtener un mayor rendimiento de la fórmula; optar por dejar preguntas en blanco tiende a bajar la nota. Es decir, la fórmula clásica de penalización de la adivinación al azar no solo no disuadiría la adivinación, sino que la incentivaría; esa es la contradicción a la que nos referimos

más arriba. Naturalmente, no todos los alumnos tienen capacidad matemática como para darse cuenta de esto, pero la experiencia suele mostrar a los alumnos que en este tipo de exámenes, sobre todo si no se está muy seguro de obtener la puntuación a la que se aspira, resulta rentable arriesgarse a contestar preguntas dudosas. Para que fuera realmente disuasoria debería descontar una cifra superior de puntos que la establecida en la fórmula clásica, o sea, debería descontar más que la fórmula: *respuestas incorrectas/(n-1)*.

La fórmula clásica de corrección de la adivinación, pese a lo dicho en el párrafo anterior, sí ejerce en la práctica un efecto psicológico disuasorio sobre la mayoría de los alumnos; pero ese efecto depende de la predisposición del alumno a asumir riesgos y es diferente en cada uno de ellos. Es decir, los alumnos con aversión al riesgo tienden a dejar más preguntas en blanco que los alumnos más osados..., y los segundos suelen resultar beneficiados en su nota por ello, mientras que los primeros salen perjudicados. La contradicción es evidente: es un procedimiento de disuasión de la adivinación que castiga a los alumnos que precisamente la evitan. Por otro lado, siendo la tendencia al riesgo un factor en parte inherente al alumno –dependiente, en parte, de su personalidad y en parte, de sus circunstancias- no se debería incluir este factor psicológico complejo en nuestra evaluación de sus competencias en una materia; al menos, no se debería hacer de forma inconsciente y a la ligera, como desgraciadamente se está haciendo. Más adelante expondremos que nosotros incluimos en nuestro sistema de evaluación (el *test basado en la confianza*) este factor, la tendencia a asumir riesgos del alumno, pero lo hacemos de forma consciente y ponderada, obteniendo, por añadidura, información relevante del proceso de aprendizaje del sujeto.

Una forma diferente de aplicar la fórmula *clásica* -diferente en apariencia, porque matemáticamente es la misma regla- de penalización por azar sería el procedimiento de descontar del total de aciertos obtenido el porcentaje que de ellos correspondería con máxima probabilidad al azar, según el número de respuestas de cada pregunta. Se descontaría del número de preguntas (N) y del número de aciertos (A) el cociente N/n (n = número de respuestas por ítem), y la nota saldría de la siguiente fórmula: Puntuación = $(A - N/n) / (N - N/n)$. Así, por ejemplo, en un test con cuatro respuestas por cada pregunta habría un 25 % de probabilidades de acertar por azar (1/4 sobre 1), del total de respuestas acertadas se descontaría la cuarta parte de las preguntas totales, y también se descontaría esa misma cuarta parte del total de las preguntas. De esa manera, en un examen de 40 preguntas con cuatro respuestas, si el alumno contestase a 25 correctamente y tuviera 15 incorrectas, su puntuación sería de 5 (o de un 50 %) $[25 - 40/4] / (40 - 40/4)$ igual a 15/30, igual a 5 sobre 10]. Si en esta situación aplicáramos la fórmula clásica de corrección de la adivinación $[(25 - 15/3) / 40]$, obtendríamos exactamente la misma puntuación. Al empezar a hacer un examen de estas características, al alumno simplemente le informamos del número de respuestas que debe acertar para aprobar, o para obtener la nota que sea, e igualmente le decimos que solo cuentan las respuestas correctas, siendo indiferente que el alumno deje una pregunta en blanco o la conteste de forma errónea. En realidad, es una forma diferente de expresión de la misma fórmula matemática clásica. Pero es importante destacar que cuando escogemos esta forma de puntuar, descontando del total de preguntas y del total de aciertos N/n , resulta muy evidente para el propio alumno que debería responder a todas las preguntas para conseguir el mayor rendimiento, supiera o no la respuesta correcta, pues dejarlas en blanco le perjudicaría lo mismo que responder de manera

incorrecta. Es decir, esta variedad de la fórmula clásica de corrección no busca disuadir de utilizarla adivinación, sino ponderarla; por ello nos parece preferible.

En conclusión, cualquier procedimiento que sigamos para disuadir al alumno de jugar con la adivinación va a distorsionar la puntuación del alumno introduciendo factores psicológicos y personales. Además, ¿para qué queremos disuadir al alumno de practicarla adivinación? Aunque, como dijimos más arriba, algunos autores razonen que la adivinación es una mala práctica en el alumno (Oh, 2004), nosotros creemos que es preferible no el evitarla sino ponderar su efecto o, incluso, utilizarla para obtener más información del aprendizaje del alumno. Cuanto más recurre el alumno en un examen de tipo test a la adivinación, menor es su confianza en obtener un buen resultado. La adivinación y la confianza son, por lo tanto, antitéticas. Y la confianza, como expondremos más adelante, es una expresión de la solidez y profundidad del conocimiento, de la *monitorización* de su propio conocimiento. En lugar de disuadir la adivinación, nosotros la utilizamos en un sistema de examen que pondera la confianza como una dimensión fundamental del aprendizaje. En nuestro sistema de examen basado en la confianza, utilizamos la adivinación para obtener un conocimiento más completo del aprendizaje del alumno.

2.1.3. INCONVENIENTES DE LA CORRECCIÓN QUE PENALIZA LAS RESPUESTAS ERRÓNEAS

El procedimiento de penalizar las respuestas erróneas tiene, además de la objeción que le hemos hecho con anterioridad, serios inconvenientes. Funcionaría correctamente si el alumno, ante una pregunta, comprendiera perfectamente el enunciado de esta y solo tuviera dos posibilidades: o saber exactamente y con seguridad la respuesta o ignorarla

absolutamente (Morales, 2006a). Pero en la práctica eso no es así; hay diversos grados de conocimiento intermedio entre el conocimiento seguro y el completo desconocimiento, que revelan grados parciales de dominio de la materia y a veces también dudas semánticas en la interpretación del enunciado de la pregunta, o variables de otro tipo, como la tendencia personal de cada estudiante a asumir riesgos, que hemos mencionado antes. De esta manera, la mayoría de los alumnos suelen estar en la siguiente situación: conocen con certeza la respuesta de algunas preguntas, desconocen totalmente otras y tienen dudas de diferente grado frente a un tercer grupo.

Un conocimiento es seguro cuando el sujeto que lo posee puede explicar de forma racional sus fundamentos, los motivos por los que algo es de una manera y no de las otras formas posibles (Morales, 2006a). Si el sujeto no es capaz de dar ninguna explicación racional para una idea o concepto, salvo la de haber leído u oído que ese concepto o idea era la correcta, el conocimiento deja de ser seguro, pues por bien que recuerde dónde leyó u oyó eso, no podrá descartar que la fuente no estuviera equivocada. En definitiva, definimos el *conocimiento de calidad* siguiendo la definición aristotélica de conocimiento cierto como «*conocimiento verdadero y justificado*» (Gardner–Medwin y Gahan, 2003). El recurso acrítico a la autoridad de las citas como base del conocimiento se abandonó precisamente cuando se comenzó a desarrollar el conocimiento científico, que de una forma u otra se basa en una cadena de razonamientos lógicos para explicar los hallazgos empíricos. Por ello, para que se pueda considerar seguro el conocimiento ha de estar necesariamente *justificado*.

En las preguntas de tres o más respuestas es frecuente que el alumno pueda descartar totalmente una o varias de ellas y dude solo entre las restantes, lo que indica ya un cierto grado de conocimiento de la materia. La penalización por respuesta incorrecta hace que los alumnos se comporten en estas situaciones de forma diferente

según su disponibilidad personal a aceptar el riesgo. Los alumnos más osados se atreverán a responder en los casos dudosos, bastándoles con poder descartar alguna de las alternativas que juzgan erróneas, y los más prudentes, o timoratos, se abstendrán de responder cuando no estén muy seguros. En ocasiones son las expectativas de nota del alumno las que condicionan el nivel de asunción de riesgos; el alumno con insuficiente dominio de la materia pero que conserva esperanzas de aprobar sabe que se tiene que arriesgar para conseguirlo, y el que ya ha asegurado la nota que esperaba obtener con las preguntas en las que está seguro, frecuentemente tenderá a ser muy cauto con las demás para no correr el riesgo de ser penalizado. Como dijimos antes, con este sistema los alumnos que aceptan riesgos son beneficiados en la mayoría de las ocasiones en sus resultados frente a los excesivamente cautelosos, pues al responder a las preguntas dudosas –pero en las que hay un conocimiento parcial- el porcentaje de aciertos suele compensar en exceso la penalización que se recibe por las respuestas equivocadas, y esto sucede especialmente en los alumnos mejor preparados. No olvidemos que, como dijimos antes, la penalización de las respuestas erróneas está calculada para el desconocimiento absoluto, para el azar puro. Esto fue puesto ya de manifiesto en 1929 por Greene, y se confirma regularmente en estudios posteriores (Morales, 2006 a).

Así, las características psicológicas del alumno y su situación personal se convierten en un condicionante de los resultados. Se ha comprobado que las alumnas en este tipo de exámenes resultan perjudicadas, por término medio, frente a los alumnos (Ben-Shakhar y Sinai, 1991; Budescu y Bar-Hillel, 1993; Hassmen y Hunt, 1994), lo que introduce un sesgo de género. A igualdad de conocimientos, la prudencia y el carácter conservador bajan la puntuación; esto entendemos que no es deseable, pues no son las características personales del alumno lo que habitualmente pretendemos valorar en nuestros exámenes.

2.1.4. ALTERNATIVAS AL MÉTODO TRADICIONAL DE CORRECCIÓN DEL AZAR EN LOS EXÁMENES DE TIPO TEST

Se han propuesto múltiples alternativas para evitar los inconvenientes que surgen del sistema de corrección de las pruebas objetivas de elección múltiple con penalización de las respuestas erróneas; además de controlar el efecto de la adivinación, varias de ellas intentan ponderar el conocimiento parcial. Las más importantes aparecen resumidas en la tesis doctoral de Belén Urosa (1995); mencionaremos las siguientes:

- A) Preguntas con múltiples respuestas correctas (*multiple-marksquestions*, MMQ) (Pomplun y Omar, 1997).
- B) Selección de alternativas incorrectas (Coombs, 1953; Coombs, Miholland y Womer, 1956; Collet, 1971).
- C) Técnica de selección del subconjunto (Dressel y Schmid, 1953; Urosa, 1995).
- D) Responder hasta la correcta (Frery, 1982; Wicox, 1982).
- E) Fórmula que premia las omisiones (Gulliksen, 1950).
- F) Fórmula que premia las omisiones y penaliza los errores (Reilly, 1975).
- G) Fórmula de Reid (Reid, 1977).
- H) Test de eliminación de las respuestas correctas (*number-right elimination testing*, en acrónimo NRET) (Chang, Ling P.C. y Ling Z.C., 2007; Lau, 2011; Lau y otros, 2014).

I) Preguntas de aseveración-razonamiento (ASQ o *assertion-reason question*; Connelly, 2004; Williams, 2006).

J) Técnica del grado de seguridad en la selección de la respuesta (*confidence based testing*, o *confidence-based marked testing*, CBT o CBM; en español «test basado en la confianza», TBC, o también «test basado en el grado de seguridad») (Gardner-Medwin 2006^a y b, 2007; Morgan y Cleave-Hogg, 2002).

El último sistema de examen de pruebas objetivas múltiples es el que estamos utilizando en nuestro centro y sobre el que desarrollamos el presente estudio. Aunque el nombre más correcto en castellano sería «test con puntuación basada en el grado de seguridad» o en la «auto-confianza» –esta última es la denominación que sería más adecuada en castellano-, al no haber un nombre específico en nuestro país, le llamaremos a partir de ahora, por simplificar y usar el término más parecido al inglés, «test basado en la confianza», TBC en siglas.

Los sistemas de examen basados en la confianza se caracterizan porque no son unidimensionales sino bidimensionales; evalúan las dos dimensiones siguientes: *conocimiento* (o competencia, según los casos) y *confianza* en que el conocimiento es correcto. Esa bidimensionalidad constituye su aportación original a los sistemas de evaluación.

Antes de estudiar los sistemas de examen en detalle trataremos la *confianza* académica, sus características y su engarce dentro del campo amplio de la *regulación del aprendizaje* y la *metacognición*.

2.2. LA CONFIANZA COMO INDICADOR DE APRENDIZAJE AUTO-REGULADO

2.2.1. APRENDIZAJE AUTORREGULADO

Para la adquisición ordenada y coherente de conocimientos y competencias en el proceso de aprendizaje, y para que estos conocimientos y competencias sean funcionales, el alumno habrá de desarrollar un proceso de autorregulación o, dicho de forma similar, un *aprendizaje autorregulado*. El uso de estrategias adecuadas autorregulatorias del aprendizaje es fundamental para el éxito académico en los tres niveles educativos, en educación primaria, secundaria y universitaria (Panadero y Alonso-Tapia, 2014; Salmerón, Gutiérrez, Salmerón y Rodríguez, 2011; Sitzmann y Ely, 2011).

La auto-regulación del aprendizaje es un concepto que se hace popular en la psicopedagogía a partir de los años 80 del siglo pasado, al enfatizar la autonomía y responsabilidad de los alumnos ante las tareas de aprendizaje (Álvarez, 2009) frente a la concepción tradicional de enseñanza normativa y dirigida. No es intención del presente trabajo profundizar en las distintas concepciones teóricas del aprendizaje autorregulado, cuestión sumamente compleja que sobrepasa los propósitos de este Marco Teórico, mucho más cuando, como señala Álvarez, existe una «multiplicidad de definiciones y concepciones de autorregulación que está presente en la investigación contemporánea (...) en la psicología de la educación», existiendo numerosas confusiones conceptuales e incluso interpretaciones diferentes de las contrastaciones empíricas (Álvarez, 2009). Sin profundizar en este complejo tema, señalaremos escuetamente las características del proceso de aprendizaje autorregulado, asumiendo la multiplicidad de enfoques que actualmente existen sobre el tema.

Se puede hablar de tres niveles de regulación en el aprendizaje: la autorregulación, la regulación externa y la ausencia de regulación (Vermunt, citado por Salmerón y otros, 2011). La primera categoría de regulación, la autorregulación, sería la que conllevaría unas actividades de aprendizaje de mayor nivel y se asociaría a mejores resultados académicos, si bien se señala por algún autor que la evidencia de la asociación entre autorregulación y resultados académicos no siempre resulta consistente (Salmerón y otros, 2011).

La autorregulación del aprendizaje para Pintrich (1995) es *«el autocontrol activo del comportamiento del estudiante dirigido a un objetivo, que exige motivación y conocimiento de las tareas académicas»*. Zimmerman la define como *«la habilidad para reconocer los propios objetivos y las estrategias necesarias para alcanzarlos, incluyendo evaluar el progreso hacia los objetivos y conocer cómo conseguir las ayudas apropiadas»* (Zimmerman 2002, pp. 1-12). Una definición más completa, basada en los modelos de Zimmerman, es la que proporcionan Panadero y Alonso-Tapia; definen la autorregulación como *«el control que los estudiantes tienen sobre su cognición, comportamiento, emociones y motivación, a través del uso de estrategias personales para conseguir los objetivos que ellos mismos han establecido»* (Panadero y otros, 2014). Tal como exponen Ziemmerman (2002), Isaacson y Fujita (2006) y Panadero y otros (2014), la autorregulación es un proceso complejo que incluye no solo la aplicación de una serie de estrategias para conseguir unos objetivos sino también una importante autorreflexión y autoconocimiento, motivación, emociones e interacción social. Los alumnos expertos y de alto rendimiento destacan por su autorregulación. Ellos no solamente tienen un arsenal de estrategias de aprendizaje adaptadas a cada tarea, y la habilidad para regular su actividad adaptándola al cumplimiento de los objetivos, también saben cuándo han cubierto los objetivos del aprendizaje, cuándo han

alcanzado la «maestría» en las tareas académicas y cuando no las han conseguido y deben continuar con su esfuerzo (Isaacson y Fujita, 2006).

Zimmerman, en sus investigaciones sobre los procesos de aprendizaje, fue desarrollando sucesivamente varios modelos social cognitivos sobre el aprendizaje autorregulado: un primer modelo «*triádico*» en el que aúna factores intrínsecos, de comportamiento y ambientales en un feedback personal; un segundo modelo multinivel de aprendizaje, que comienza con un aprendizaje observacional y de emulación y evoluciona secuencialmente hasta la autorregulación, y, finalmente, su modelo cíclico, que el autor desarrolla en años recientes (Zimmerman, 2013). Vamos a describir, con algo más de detalle, este último modelo.

En el modelo cíclico, Zimmerman establece tres fases en el proceso de autorregulación (Panadero y otros, 2014; Zimmerman, 2013; Zimmerman y Tsikalas, 2005):

- 1) La primera fase, fase de planificación, o «premeditación», es la del establecimiento de objetivos y de la planificación estratégica para conseguirlos; no requiere un esfuerzo de voluntad pero sí análisis de las tareas a realizar y un conjunto de creencias auto-motivantes. El análisis de las tareas requiere el *establecimiento de objetivos* y la *planificación estratégica*. Las creencias automotivantes son la *autoeficacia*, las *expectativas de éxito*, el *interés/utilidad* de las tareas y *focalización hacia el objetivo*.
- 2) La segunda fase sería la fase de «desempeño», desarrollo y ejecución de las tareas planificadas en la primera. Tiene dos grandes componentes:

autocontrol y auto-observación. El autocontrol requiere un esfuerzo voluntario que incluye: *estrategia* para las tareas, *auto-instrucciones*, *control del tiempo*, *estructuración del entorno*, *incentivos* y *autoconsecuencias* y *búsqueda de ayudas*. La auto-observación requiere de *monitorización metacognitiva* y puede potenciarse con el recuerdo de procesos anteriores.

- 3) La tercera sería la fase de autorreflexión, que compara los resultados que se van obteniendo, y que se conocen mediante la *automonitorización*, con los objetivos del aprendizaje; esta tercera fase sería la *autoevaluación* de los resultados y la *atribución de causas* a éstos. También, en el modelo cíclico de Zimmerman, en esta tercera fase se incluye la reacción del estudiante –*satisfacción* o *insatisfacción*- y su respuesta –*adaptativa* o *defensiva*-.
- 4) Tras la autoevaluación de la tercera fase, los estudiantes, pasan de nuevo a la fase inicial para replantearse sus objetivos y estrategias, lo que cierra el ciclo.

En estas tres fases destacan tres factores clave: el establecimiento y seguimiento de **objetivos**, la **motivación** y la **auto-monitorización** –también llamada *monitorización metacognitiva*-. Los estudiantes utilizan los **objetivos** de aprendizaje para planificar la estrategia y para auto-motivarse en la primera fase, para mantener el incentivo e interés en la tarea en la fase de desempeño y, finalmente, como la vara de medir sus resultados, en la última fase del proceso. Igualmente relevantes son la **motivación** y la **monitorización metacognitiva** en las todas las fases del proceso

(Panadero y otros, 2014). Es importante destacar, como hacen Ramdass y Zimmerman (2011), que ese conjunto de habilidades autorregulatorias que fueron desarrolladas durante la etapa de aprendizaje del individuo, como *establecimiento de objetivos, manejo del tiempo y automonitorización*, no solamente son importantes para el éxito académico, sino que se acaban convirtiendo en componentes fundamentales de la vida de profesionales, deportistas, artistas y científicos de éxito.

Como acabamos de ver, de acuerdo con la última concepción de Zimmerman, el proceso de aprendizaje autorregulado es un proceso activo, cíclico, recurrente, y motivado (Álvarez, 2009). Se han señalado algunas carencias en este modelo: no incluye prácticamente los aspectos emotivos –solo, en parte, en la tercera fase- y muy poco los aspectos sociales. En otros modelos, como en el modelo volicional de Kuhl y en el Bockaerts (Panadero y otros, 2014), las emociones del estudiante son parte fundamental y su control es necesario para el éxito. En cuanto a los aspectos sociales, poco mencionados en el modelo cíclico de autorregulación, se ha señalado acertadamente que el desarrollo de la competencia autorreguladora no es el resultado exclusivamente de un proceso intrínseco –la maduración del estudiante- ni tampoco es solamente un producto extrínseco –emergente a partir del ambiente-, es un producto de la integración de ambos factores, extrínseco en intrínseco (Salmerón y Gutiérrez-Braojos, 2012). En este sentido, previamente a la elaboración de su modelo cíclico, Zimmerman elaboró un modelo explicativo del desarrollo de la autorregulación con cuatro niveles de desarrollo –observación, emulación, autocontrol y autorregulación- de los cuales los dos primeros tienen una gran influencia social, mientras que los dos segundos son, fundamentalmente, intrínsecos (Salmerón y otros, 2012).

En el proceso de autorregulación, un elemento fundamental es la capacidad del estudiante para monitorizar su aprendizaje en todas sus fases, tanto en la fase inicial, de establecimiento de objetivos y estrategias, como en la de acción y de evaluación; la autorreflexión y la automonitorización son elementos críticos a lo largo de todo el proceso; los alumnos conscientes de su nivel de maestría pueden ajustar sus estrategias y prácticas de estudio en función de los resultados parciales que van detectando. La adecuación de esta monitorización del desempeño del estudiante es crucial. Cuando el alumno no es adecuadamente consciente de su nivel de maestría puede caer en las dos situaciones opuestas: sobreestimar o infraestimar su nivel de competencia académica (Kelemen, 2000). Los alumnos que tienden a sobreestimar su conocimiento suelen finalizar su esfuerzo de aprendizaje cuando aún no se han alcanzado las metas; luego, cuando los resultados de los exámenes o evaluaciones les muestran su pobre desempeño, suelen reaccionar con sorpresa o incredulidad. Los profesores estamos acostumbrados a esta situación; gran parte de estos alumnos no serán conscientes de su deficiente esfuerzo de preparación y atribuirán su resultado, inferior al esperado, a una excesiva dificultad del examen, a la animadversión del profesor o a la mala suerte. Por otro lado, los alumnos que infraestiman su competencia académica, a veces por un pobre autoconcepto y otras por una mala estimación de las metas del aprendizaje de esa materia concreta, tienden a manifestar un defecto en la autoconfianza (Fong y Krause, 2014); esa falta de confianza no les llevará a adquirir una mayor maestría, sino que interferirá negativamente con el desempeño futuro de sus tareas en esa materia. La monitorización correcta y adecuada del conocimiento aparece como un punto crucial en el aprendizaje; los alumnos consiguen tener éxito en la medida en que tienen objetivos propios claros y se mantienen informados de su progresión hacia esos objetivos mediante una constante **automonitorización** (Fenning y May, 2013).

Para una correcta autorregulación, es fundamental que el alumno posea unas adecuadas *motivación*, *metacognición* (o «monitorización metacognitiva») y *autoeficacia*. A partir de la motivación, del conocimiento de los propios conocimientos, capacidades y de las características de la tarea que aporta la metacognición, y de la correcta estimación del desarrollo futuro de las capacidades de aprendizaje, que supone la autoeficacia, se pueden establecer objetivos realistas y ejercer el control sobre el desarrollo de todo el proceso.

En este ciclo de autorregulación del aprendizaje, la *metacognición*, desarrollada mediante un proceso de autorreflexión y análisis que llamamos automonitorización, es un requisito para poder llevar a cabo la planificación y evaluar permanentemente el desempeño de las tareas. En el pasado se tendía a considerar a la metacognición como único proceso en la fase precursora del aprendizaje estratégico (Boekaerts y Corno, 2005); actualmente suele ser vista, más bien, como un componente más de la autorregulación (Panadero y otros, 2014). Por otro lado, el concepto de metacognición es sumamente complejo también y, al igual que habíamos dicho sobre el *aprendizaje autorregulado*, el análisis detallado de la abundantísima literatura sobre él excede con mucho de los objetivos y extensión del presente Marco Teórico; no obstante, como el sistema de examen que estudiamos se basa en la reflexión del alumno sobre su propio nivel de conocimientos y la confianza que surge como consecuencia, en definitiva sobre su metacognición, hablaremos, si bien brevemente, de ella.

2.2.1.1. METACOGNICIÓN, AUTOMONITORIZACIÓN

Como señalamos antes, la autorregulación precisa entre sus componentes de un *metaconocimiento* -o *metacognición*- (Winne, 1996); conocimiento tanto sobre los

propios procesos y productos cognitivos como sobre las propiedades de la información y los otros datos relevantes para el aprendizaje (González, García y Vázquez, 2014). Definimos a la metacognición como la competencia que permite ser conscientes de nuestra propia cognición, al menos de parte de ella (Pérez de Albéniz, Escolano, Pascual, Lucas y Sastre, 2015). Partiendo de esa base metacognitiva, el estudiante podrá desarrollar el ciclo continuo de aprendizaje autorregulado.

La metacognición es el conocimiento que el sujeto tiene de sus propios campos cognitivos, es decir, el *conocimiento y conciencia del propio conocimiento* (Pintrich, 2002). Fue definida inicialmente como el conocimiento y la regulación de las propias actividades cognitivas en el proceso de aprendizaje (Brown, 1978, página 68; Flavell, 1979). Es un concepto que se empieza a utilizar en el campo de la psicopedagogía, a medida que se intenta profundizar en los procesos psiconeurológicos que determinan el aprendizaje. Se suele concebir como un conocimiento de alto nivel sobre el propio conocimiento (Brown, 1978; Veenman, 2006); es un agente que sobrevuela y gobierna todo el campo cognitivo, del que, a su vez, forma parte (Veenman, 2006); un conocimiento exacto de lo que sabe y de lo que puede hacer con lo que sabe, el *conocimiento consciente de sí mismo*. Juega un importante papel en todos los procesos de aprendizaje, ya desde la infancia, en la comunicación oral de información, la comprensión oral y lectora, la escritura y adquisición del lenguaje, la memoria, la capacidad de resolver problemas y el autocontrol y autoinstrucciones (Flavell, 1979). Wang, Haertel y Walberg, en 1990, en un extenso meta-análisis de todo lo publicado hasta entonces, llegaron a la conclusión de que entre las más significativas influencias en el aprendizaje, los factores metacognitivos ocupaban la segunda posición en importancia, después -pero a corta distancia- de la capacidad docente del profesor en el aula. El estudio sobre los aspectos metacognitivos del aprendizaje, la monitorización y

el control sobre los conocimientos, puede producir sinergias entre los enfoques psicológicos y filosóficos sobre la consciencia (Nelson, 1996).

Flavell (1979) construyó el primer modelo sobre la metacognición, que consideraba cuatro componentes:

- 1- El conocimiento «metacognitivo», a su vez dividido en conocimiento sobre la persona, sobre la tarea a realizar y sobre las estrategias posibles.
- 2- La experiencia de procesos anteriores («experiencia metacognitiva»).
- 3- Las metas a conseguir.
- 4- Las estrategias a implementar.

Este modelo anticipa, de alguna manera, lo que en Zimmerman sería el ciclo de la autorregulación en su fase inicial. La metacognición constituiría, así, el prerequisite fundamental para el aprendizaje autorregulado, que se llevaría a cabo implementando a continuación las estrategias y acciones diseñadas y evaluando los resultados, como dijimos antes.

La adecuada metacognición en un campo exige un adecuado dominio de dicho campo; como dice Veenman (2006), «la metacognición se nutre de la cognición». Sin conocer lo que es relevante e irrelevante, las dificultades intrínsecas del aprendizaje de un campo cognitivo, el sujeto es incapaz de juzgar sobre sus competencias en el mismo. El objetivo no puede ser solamente *saber* o *saber hacer* sino que debe *conocer lo que se sabe*, y, en consecuencia, lo que se ignora o no se puede hacer, tanto en conocimientos teóricos como en competencias prácticas. El aprendizaje no se puede regular si no es a

través de un proceso metacognitivo (Winne, 1996), y en ausencia de este, el proceso de aprendizaje resultaría inseguro y, necesariamente, incompleto. En este sentido, se ha considerado que la incapacidad de controlar o manejar los propios procesos cognitivos es parcialmente responsable del bajo rendimiento en las tareas de aprendizaje (Reeve y Brown, 1984).

Dentro del término *metacognición*, nos referimos a varios procesos, que Martí (1995) y Pintrich (2002) resumen básicamente en dos:

1.- *El conocimiento sobre los procesos cognitivos*: incluye la cognición de los propios conocimientos ya fijados, de las lagunas de conocimiento, de las necesidades de conocimiento futuras, de las capacidades propias de aprendizaje y de las características y exigencias que tendrán las tareas futuras de aprendizaje (por ejemplo, saber que no se ha comprendido algo o, ante una serie de tareas futuras, saber cuál es la más fácil o las difíciles). Para este primer proceso es necesaria un *escrutinio* o «automonitorización» del conocimiento (Sharma y Bewes 2011), así como un previo conocimiento de las capacidades, habitualmente incluido dentro del amplio término de *autoconcepto*.

2.- *La regulación de los procesos cognitivos*. Constituye un grupo de procesos que incluyen la *monitorización, control y regulación* de la propia cognición -«área de control»- (Pintrich, 2002). Este segundo proceso entronca ya con el concepto de autorregulación, que vimos antes.

En el segundo aspecto, la regulación del aprendizaje, Pintrich distingue tres tareas o procesos: primero, **el conocimiento general de las estrategias** de aprendizaje («conocimiento estratégico»); segundo, **conocer y saber cuándo utilizar las tareas**

cognitivas necesarias en el contexto del estudiante, y tercero, el **autoconocimiento**, referido tanto a los factores cognitivos como motivacionales (Pintrich, 2002). En cuanto al segundo factor –conocimiento y competencia del estudiante en las tareas cognitivas necesarias en su contexto- Anne Brown (1987) distingue tres tareas específicas: *planificación*, antes del inicio de las tareas; *control*, durante el desarrollo de estas; *evaluación* de los resultados, al finalizar.

- 1) *Planificación*: En la primera fase, los sujetos identifican la tarea, generan objetivos y adoptan un plan para alcanzar dichos objetivos. En esa fase inicial, partiendo de su motivación para hacer la tarea, han de examinar la masa de información que juzgan relevante, sus capacidades personales y prever los obstáculos con que se pueden encontrar y los recursos con los que pueden contar (Winne, 1996). Crearán con ello una estrategia general y una serie de tácticas a aplicar, en definitiva, un plan de actuación para el aprendizaje.
- 2) *Control*: En la segunda fase, el estudiante irá juzgando la idoneidad de la estrategia y las tácticas que va empleando, así como la adecuación de la planificación inicial, y, de acuerdo con los resultados parciales y la nueva información que va adquiriendo, juzgará si la estrategia general y las tácticas eran las adecuadas y, si lo ve necesario, las modificará. Todo ello constituye un proceso continuo de regulación y retroalimentación, que culmina en la tercera fase o fase de *evaluación*.
- 3) *Evaluación*: La autoevaluación de resultados será la que le permitirá al estudiante prever cuáles serán sus resultados académicos, en el examen o en otros procedimientos evaluativos del profesor, y modificar sus estrategias de aprendizaje si es necesario para obtener el resultado deseado (Winne, 1996).

Como se puede ver, el modelo de procesos metacognitivos de Anne Brown es en cierta manera semejante, paralelo, al ciclo de autorregulación de Zimmerman. Podríamos decir que Anne Brown está representando en su modelo lo que Zimmerman denominaría la *monitorización metacognitiva*, acción que se desarrolla en las tres fases del ciclo autorregulatorio, y, en consecuencia está anticipando la definición de ese ciclo, al menos en su aspecto puramente *cognitivo* (Brown, 1987, Zimmerman 2013, Panadero y otros, 2015).

Una forma más sencilla de definir la metacognición, que parte también de los postulados anteriores, establece dos componentes de ésta: el conocimiento sobre los propios procesos y productos cognitivos –autoconocimiento– y el conocimiento sobre las propiedades de la información y los demás datos relevantes, externos, para el aprendizaje (Flavell, 1979; González, García y Vázquez, 2014).

La metacognición es un componente fundamental del aprendizaje autorregulado. Los alumnos capaces de autorregularse son capaces de establecer qué deben saber, qué no saben y qué conceptos deben comprender; se automonitorizan, analizan su propio desempeño, lo evalúan y actúan en consecuencia (González, García y Vázquez, 2014). Muchos autores consideran que la autoconfianza es el resultado, y también la medida más precisa, de la «automonitorización» en sus diversas tareas (Sharma y Bewes, 2011).

- **Conocimiento y metacognición.**

La *metacognición*, como dijimos antes, «se nutre de la cognición» (Veeman, 2006), forma parte del mismo proceso y no siempre es fácil separarlas nítidamente. A esta dificultad para deslindar ambos conceptos, se une un cierto abuso en las últimas décadas

de los términos psicológicos y pedagógicos encabezados por el prefijo «meta» Como dijo Eduardo Martí: «Esta fascinación por lo “meta” ha conducido (...) a un cierto desaliento: a medida que el prefijo se ha ido extendiendo de forma irrefrenable, aplicándose prácticamente a cualquier proceso cognitivo (metamemoria, meta-aprendizaje, metacompreensión, meta-atención, metarrepresentación, meta-imitación, metaplanificación, metaestrategias, etc.), el significado del término “metacognitivo” ha ido desdibujándose y la pertinencia de las investigaciones metacognitivas ha llegado a ponerse en entredicho» (Martí, 1995).

Admitiendo el abuso de los conceptos psicopedagógicos encabezados por «meta» y la dificultad en separar nítidamente conocimiento y metaconocimiento, el control que la metacognición supone del campo cognitivo (lo «automonitoriza», cuantificándolo y detectando sus lagunas, lo relaciona y asocia con los demás conocimientos, encuadrándolo en un marco lógico y funcional) es necesario para la correcta aplicación práctica de ese conocimiento. La metacognición no solo sobrevuela, conoce y controla el conocimiento, también permite la «praxis», que sin esa función de control metacognitiva resultaría insegura y titubeante. En los términos habituales de las neurociencias, si la falta de conocimiento genera «agnosia», la falta de metacognición generaría «apraxia», haciendo inútil el propio conocimiento.

Dentro del ancho campo de la metacognición aparece un grupo de subcampos, o componentes que no siempre resultan fáciles de sistematizar. Los conceptos más utilizados son: la *automonitorización*, que es el proceso de análisis de la situación cognitiva en el momento presente; el *autoconcepto*, el conocimiento que uno tiene de sí mismo, un análisis de uno mismo que mira al pasado y tiende a ser estable o cambiar lentamente (Marsh y Craven, 2006); la propia *autoeficacia* (de la que hablaremos a

continuación), concepto desarrollado por Bandura (1977, 1986, 1993), referido a la capacidad autopercebida de poder realizar las tareas exigidas para conseguir un objetivo, y la *confianza*, o *autoconfianza*, la convicción de ser capaz de desempeñar con éxito una tarea concreta.

La multiplicación de términos en torno a estos conceptos psicológicos, los señalados antes, y otros muchos como *expectativas de rendimiento*, *autopercepciones de capacidad*, *competencia percibida*, *control autopercebido*, etc. (Pajares, 2003), con diferencias entre ellos demasiado sutiles -todo ello empeorado a veces por la traducción de esos términos del inglés a otras lenguas-, puede crear una notable confusión en quien se aproxima por primeras veces a estos términos psicopedagógicos, e incluso en ocasiones en el experto (Torre, 2007, pp. 41-43). Una forma inicial de intentar ordenarlos y clarificarlos es considerar que son aspectos parciales de la metacognición y, como tales, prerequisites para un aprendizaje autorregulado.

2.2.1.2. AUTOEFICACIA

El concepto de autoeficacia se populariza a partir del trabajo germinal de Bandura (1977). La autoeficacia académica es la convicción que el estudiante tiene en que podrá ejecutar exitosamente el comportamiento requerido para conseguir sus metas de aprendizaje; Bandura (1977, 1993) la define como el «*juicio sobre la propia capacidad y habilidad para organizar y ejecutar determinados tipos de rendimientos*». Es un concepto muy próximo al de *autoconfianza*, aunque sutilmente diferente, como luego expondremos. Las fuentes de la autoeficacia, lo que hace que las personas confíen en

ser capaces de realizar las acciones necesarias para conseguir sus metas, son, en la concepción de Bandura, cuatro (Bandura 1986; Usher y Pajares 2006, 2008):

- 1) La *experiencia de maestría*, que es la experiencia del estudiante de haber tenido éxito previamente en tareas parecidas.
- 2) La *persuasión verbal*, el apoyo y estímulo verbal que otras personas realicen, muy especialmente los docentes.
- 3) La *retroalimentación indirecta o experiencias vicarias*. Con experiencias vicarias se entiende la observación de la experiencia de los otros en esas mismas tareas, especialmente si es exitosa.
- 4) La *retroalimentación fisiológica y emocional*. Este término se refiere fundamentalmente a la autopercepción de estrés u otras sensaciones fisiológicas que el estudiante tiene ante la tarea y durante la misma.

Para Bandura (1986), el más importante de estos cuatro factores es el primero, la experiencia de éxito en tareas similares o parecidas. Los éxitos previos refuerzan la autoeficacia del estudiante, los fracasos la disminuyen. Fong y Krause (2014) resaltan, además de la experiencia de maestría, la persuasión verbal; especialmente la retroalimentación positiva de los docentes es un factor fundamental en el éxito académico.

- **Autoeficacia como predictor de éxito académico.**

La autoeficacia del estudiante tiene valor predictivo de sus resultados académicos, los estudiantes con altos niveles de autoeficacia establecen metas más altas, usan

estrategias autorregulatorias más efectivas, se automonitorizan más eficientemente, perseveran más cuando hacen frente a tareas académicas cambiantes y evalúan de forma más ajustada sus resultados comparados con los estudiantes con baja autoeficacia (Ramdass y otros, 2008). Pero algunos autores llaman la atención sobre el hecho de que la autoeficacia, de alguna manera, predice los resultados académicos como una «*profecía autocumplida*» (Multon y otros, 1991). Este es un hecho muy conocido por todos los que han tenido experiencias docentes; aumentar la autoeficacia académica proporcionando retroalimentación positiva de los resultados, y limitar el efecto negativo que los fracasos ejercen en la confianza del alumno, cuando ello es posible, es una táctica eficaz para mejorar el rendimiento del estudiante y utilizada por los profesores hábiles (Fong y Krause, 2014; Ramdass y otros, 2008).

Se tiende a confundir los términos *autoeficacia* y *autoconcepto*. Es evidente la proximidad de ambos términos, pero no son lo mismo. Operan en dos dimensiones diferentes. El *autoconcepto* se refiere a la totalidad del conocimiento que uno tiene de sí mismo e incluye sentimientos de autoestima; es una autoevaluación tanto cognitiva como afectiva (Bong y Skaalvik, 2003). Torre lo define como «*un juicio autodescriptivo que incluye una valoración de la competencia y sentimientos sobre la valía personal*», la autoeficacia es, para el mismo autor «*un juicio sobre la capacidad propia para realizar una tarea o comprometerse en una actividad determinada*» (Torre 2007, pp. 44-45). Se puede apreciar que el *autoconcepto* actúa en un plano más general y conlleva un juicio global sobre la propia valía, un juicio afectivamente no neutro; es, igualmente, retrospectivo y orientado al pasado; la autoeficacia actúa en un plano mucho más concreto y delimitado; es, en cierta manera, más objetiva, y es una capacidad prospectiva, orientada al futuro. Resumiéndolo en dos frases, el *autoconcepto*

responde a la siguiente pregunta «¿qué soy yo?»; la autoeficacia respondería a esta otra pregunta: «¿puedo conseguir esto?».

2.2.1.3. CONFIANZA ACADÉMICA

En nuestro trabajo utilizamos extensamente el concepto de *confianza* (autoconfianza académica), que es la base de nuestro sistema de evaluación. Con el término *confianza* nos referimos a un hecho muy concreto: la confianza del alumno, durante el examen, de que sus conocimientos sean los adecuados para superarlo y, en consecuencia, de que sus respuestas a las preguntas sean las correctas. Otra forma de denominar a este concepto es el de *seguridad* o *grado de seguridad* en los propios conocimientos.

La *autoconfianza* o *confianza* académica tiende a equipararse a la *autoeficacia*. Aunque se ha extendido una cierta confusión entre estos dos conceptos, nosotros entendemos la confianza como una percepción mucho más centrada en actividades concretas, incluso muy concretas, delimitadas y fácilmente verificables, mientras que la autoeficacia opera en un plano menos concreto y más general. Por ejemplo, mientras que la siguiente pregunta: «¿Podrás conseguir la preparación adecuada para obtener tus metas académicas?» es una pregunta sobre la autoeficacia, esta otra: «¿Podrás asistir a las tutorías?» sería una cuestión de autoconfianza –esta última pregunta está extraída de un cuestionario de confianza académica, el ABC de Sander y Sanders (2003)-. Se puede apreciar que la distinción entre ambos términos es convencional y no hay un límite preciso y bien delimitado, pero destaca la mayor concreción, centrada en tareas muy específicas, y a veces totalmente puntuales, de la confianza.

Un ejemplo de las diferencias y similitudes entre autoconcepto, autoeficacia y confianza la encontramos en el siguiente cuadro, extraído de Sander y Sanders (2006):

Tabla 2.3. Diferencia entre autoeficacia, autoconcepto y autoconfianza

<i>Dimensiones</i>	AUTOCONCEPTO	AUTOEFICACIA	CONFIANZA
Definición	Conocimiento y percepción acerca de uno mismo	Convicción de que se pueden desempeñar determinadas tareas académicas	Confianza en la propia habilidad para engranarse en el funcionamiento que se puede requerir durante la carrera académica
Elemento central	Competencia percibida	Confianza percibida	Confianza en las propias habilidades
Composición	Estimación cognitiva y afectiva de uno mismo	Estimación cognitiva de uno mismo	Evaluación del repertorio potencial de propio funcionamiento
Naturaleza de la evaluación	Normativa e ipsativa	Referida a objetivos y normativa	Respuesta a las demandas de cada situación
Estructura	jerárquica	Cuasi jerárquica	Plana y aditiva
Orientación en tiempo	Orientada al pasado	Orientada al futuro	Orientada al futuro (puede ser también retrospectiva)
Estabilidad en el tiempo	Estable	Maleable	Maleable
Resultados predictivos	Motivación, emoción y desempeño	Motivación, emoción, conocimiento, procesos auto-regulatorios y desempeño	Motivación, reproducción, búsqueda de ayudas y desempeño

Nota. Adaptado de Bong y Skaavik, (2003) y de Sander y otros (2006).

Como se puede apreciar en la tabla 2.3, las diferencias entre *autoconcepto* con los otros dos constructos son evidentes; pero la diferencia entre *autoeficacia* y *confianza* académica es mucho más sutil. Se utiliza el término *confianza* al referirse a habilidades y capacidades muy concretas que hacen referencia a las actividades desglosadas en los aprendizajes, mientras que la *autoeficacia* es una evaluación más global de la confianza en obtener un buen desempeño académico o profesional. Mientras que la autoeficacia se refiere a los objetivos del curso, es cuasi jerárquica y se basa en la estimación cognitiva general de uno mismo (Sander y otros, 2006), la confianza académica se centra en las demandas que surgen en cada situación, no es jerárquica, sino plana y -lo hemos

destacado en negrita- es *aditiva* o sumativa. Esta última es una característica que se destaca porque nos puede ayudar a aclarar la diferencia. Desglosada la competencia en un número determinado de demandas concretas, que puede ser todo lo alto que se quiera, el alumno analiza si podrá responder a cada una, irá sumando las respuestas positivas y la suma conformará su autoconfianza académica. Por eso sus resultados predictivos se basan en la reproducción de las demandas que el aprendizaje le supone, en la búsqueda de ayudas y en la estimación de cuál será su desempeño concreto en cada una de ellas (Sander y otros, 2006).

Otra característica que puede diferenciar la autoeficacia de la confianza es su carácter retrospectivo o prospectivo, orientado al pasado o al futuro. Mientras que la autoeficacia se refiere *–por definición–* siempre al futuro, la confianza se suele referir al futuro, pero a veces se refiere a un hecho ya realizado pero pendiente de evaluación, podríamos decir una situación intermedia entre retrospectiva y prospectiva. Algunas encuestas de confianza consisten en preguntarle al alumno su opinión sobre su acierto en pruebas o actividades ya realizadas pero aún no evaluadas -como hacemos nosotros en nuestro examen de tipo test-; en este caso, la pregunta es sobre una actividad ya realizada, por lo que sería retrospectiva, aunque haya sido realizada en el pasado más próximo posible al presente, y el alumno desconoce el resultado e intenta adivinarlo, lo que, a su vez, sería prospectivo (Stankov, Lee, Luo y Hogan, 2012; Morony, Kleitman, Lee y Stankov, 2013). En cualquier caso, la confianza se asienta en un punto temporal generalmente más próximo al momento de la encuesta que la autoeficacia, refiriéndose o bien a un futuro próximo, al momento presente o, incluso a una actividad ya realizada.

Para ver más gráficamente qué se suele entender por *confianza académica*, y también las dificultades que presenta su diferenciación del concepto «*autoeficacia*»,

ponemos a continuación una encuesta de competencia, el ABC, o «Academic Behavioural Confidence Scale», una encuesta diseñada por Sander y Sanders (2003), y que es actualmente muy utilizada. Para mayor claridad, hemos traducido los enunciados al castellano, (el original lo hemos incluido en el anexo B).

Tabla 2.4.

ABC TEST

¿Hasta qué punto confías en que serás capaz de...?:	
1. Estudiar efectivamente por ti mismo en tu entorno privado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
2. Producir tu mejor trabajo bajo condiciones de examen.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
3. Responder a preguntas de un profesor ante un tribunal de profesores.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
4. Llevar adelante tu trabajo hasta conseguir tus metas del curso.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
5. Hacer una presentación ante un pequeño grupo de compañeros.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
6. Asistir a la mayoría de las clases.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
7. Conseguir buenos resultados en tu trabajo.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
8. Acoplarte bien a un productivo debate académico con tus pares.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
9. Hacer al profesor preguntas acerca de las materias que está enseñando en una tutoría.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
10. Hacer al profesor preguntas acerca de las materias que está enseñando durante una clase.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
11. Entender la materia impartida y discutida contigo por los profesores.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
12. Seguir los temas y debates en las clases.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
13. Prepararte exhaustivamente para tutoriales.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
14. Leer los materiales de preparación recomendados.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
15. Desarrollar el trabajo del curso cumpliendo los estándares.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
16. Escribir en un estilo académico apropiado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
17. Pedir ayuda si no entiendes algo.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
18. Ir al día en las lecturas.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
19. Aprovechar la mayoría de las oportunidades de estudio en el grado universitario.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
20. Aprobar las evaluaciones a la primera.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado

21. Planificar adecuados programas de revisión.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
22. Permanecer siempre suficientemente motivado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
23. Hacer tu mejor trabajo en los deberes.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
24. Asistir a las tutorías.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado

Nota. Traducción del original en inglés. Test de confianza académica ABC de Sander y Sanders (2003; 2006).

Se puede apreciar la voluntad de quienes diseñaron la encuesta ABC de desglosar las actividades concretas durante el curso académico en todas las tareas imaginables. La mayoría de las cuestiones tienden a ser muy específicas y no obligan al alumno a un gran esfuerzo de introspección para responder a ellas: «ir al día en las lecturas», «asistir a las tutorías», «pedir ayuda si no entiendes algo», etc.; algunas veces, las cuestiones podrían ser calificadas con más exactitud como ítems de autoeficacia: «Aprobar las evaluaciones a la primera», «llevar adelante tu trabajo hasta conseguir las metas del curso».

Resumiendo lo anterior, cuando la autoevaluación a la que la pregunta obliga al alumno es global, le fuerza a una revisión profunda de su *autoconcepto*, y hace referencia a metas del proceso de aprendizaje, podríamos hablar de *autoeficacia académica*; cuando el alumno solo ha de evaluar su capacidad para reproducir una conducta concreta, utilizamos el término *confianza*.

Otro ejemplo de escala de confianza académica es la de Telbys de 2010 (que reproducimos en el anexo C), una subescala de 4 ítems que hace referencia más bien a lo que podríamos llamar «autoconocimiento», si no temiéramos complicar más aún la cuestión; dicha subescala es retrospectiva, hace balance de capacidades ya conseguidas, y no mira al futuro (Telbys, 2010). No todos los autores comparten esta diferencia

conceptual entre autoeficacia y competencia académica, y suelen confundirse los términos o, en ocasiones, utilizarse indistintamente, según la escuela psicopedagógica de cada uno y la moda terminológica del momento (Torre, 2007).

En nuestra opinión, la forma más eficiente de medir la confianza consiste en utilizar los exámenes o evaluaciones para formular las encuestas, tal como hace el grupo de Morony y Stankov con el examen PISA en el campo de las matemáticas (Morony y otros, 2013). El trabajo al que hacemos referencia es solamente una encuesta de opinión, que no influye en la nota que el estudiante obtendrá; los que utilizamos el sistema de examen basado en la confianza, vamos un paso más allá: la confianza condiciona la puntuación del examen. Volveremos sobre el trabajo de Morony y colaboradores más adelante.

Como hemos visto, para que cualquier aprendizaje sea útil no es suficiente con que el alumno aprehenda los conceptos, debe ser capaz de realizar un proceso autorreflexivo de revisión del conocimiento o habilidades adquiridas, proceso que se denominan «automonitorización» o «monitorización metacognitiva», y, como consecuencia de él, adquirir la confianza en su capacidad que le permita llevar a cabo las actividades que se derivan de su aprendizaje (Morony y otros, 2013). A esto lo podríamos denominar «maestría»; la confianza es, de esta manera, una excelente medida de la «maestría».

Tradicionalmente, en la práctica docente no se dedica suficiente atención a la confianza, o seguridad en la propia competencia, como uno de los elementos fundamentales integrantes del proceso educativo. Las distintas pruebas teóricas y prácticas a las que sometemos a los estudiantes suelen medir sus conocimientos o sus

competencias adquiridas, pero no se suele tener en cuenta de forma explícita en la mayoría de los procedimientos evaluativos si esos conocimientos y/o capacidades del alumno son seguros y bien fundados o si son inseguros y azarosos. Esto es aún más evidente en la mayoría de los exámenes de tipo test, en los que la respuesta acertada recibe la misma puntuación, sea cual sea la confianza en sus conocimientos del alumno. La autoconfianza es la expresión de la firmeza y seguridad de sus conocimientos. La confianza del alumno en sus conocimientos depende de su capacidad de análisis interno de los fundamentos racionales de sus conocimientos (Veenman, 2006) o automonitorización. De esa manera la automonitorización a que obliga la pregunta sobre el grado de seguridad o confianza, permite mejorar el conocimiento, o, dicho de otra manera, un adecuado nivel de metacognición permite mejorar las limitaciones del conocimiento (Veenman, 2005).

Como hemos visto, la *confianza*, o *grado de seguridad* está estrechamente conectada con varios de los conceptos que hemos manejado antes, fundamentalmente con el de *metacognición* y *automonitorización*. El test basado en la confianza –o test basado en el grado de seguridad- se basa precisamente en el uso de las capacidades metacognitivas y de automonitorización del alumno, para obtener una idea más completa de su grado de dominio de la materia, de la que daría el simple recuento de preguntas acertadas y falladas. Además, el sistema de examen basado en la confianza se propone como sistema de autoevaluación del estudiante, para monitorizar su progreso; en esta función, le obliga a una auto-reflexión sobre su conocimiento en mayor medida que los sistemas tradicionales. Utilizado para la autoevaluación, el examen basado en el grado de seguridad se puede incardinar así en las actividades de autorregulación del alumno, tal como las concibe Zimmerman en su modelo cíclico; sería una excelente forma de realizar las tareas de automonitorización. Por todo ello, por su relación con la

metacognición, con la autoeficacia y con la autorregulación, hemos dedicado espacio en este capítulo de nuestro trabajo a esos tres aspectos de la psicopedagogía.

2.2.2. LOS EXÁMENES Y LA EVALUACIÓN DE LA CONFIANZA

La confianza, como dijimos antes, es un componente del proceso de aprendizaje que está siendo objeto de gran interés en los últimos años, añadiéndose a los anteriores constructos de «autoeficacia», «autoconcepto» y «estrés académico», como un predictor importante de resultados de aprendizaje. El grupo de investigadores en psicología educativa formado en torno a Lazar Stankov, en Australia y Sudeste Asiático, ha trabajado intensamente sobre el tema (Kleitman y Stankov, 2001; Morony y otros, 2013; Stankov y Crawford, 1997; Stankov y otros, 2012). En estos estudios se suelen utilizar o bien encuestas de confianza general, como la escala de confianza académica ABC de Sander y Sanders, expuesta en la tabla 2.6, que es una encuesta de confianza «orientada al futuro» (Sander y Sanders, 2003), o bien una encuesta relacionada con un examen o prueba que el alumno está realizando, y en la que, tras cada cuestión, se le pregunta al participante por su confianza en haber respondido bien; en este caso la confianza sería *cuasi retrospectiva*. Este último procedimiento para evaluar la confianza, que sigue una sistemática similar a la de nuestro *test basado en la confianza* - salvo que la respuesta del alumno a la pregunta de confianza no condiciona su nota- se ha utilizado para un extenso análisis de las variables no cognitivas asociadas al aprendizaje, como hemos mencionado en el capítulo anterior. Los autores (Morony y otros, 2013) realizaron su estudio sobre los resultados del examen PISA 2009 en matemáticas, sobre un total de 7167 estudiantes pertenecientes a 10 países de Europa y Asia. Los resultados de este estudio mostraron que la confianza, medida de esta manera,

era un constructo más robusto, fiable y mejor predictor de los resultados académicos que la autoeficacia, el autoconcepto y la ansiedad, que también son analizados en esos países en el estudio PISA desde el año 2003. Los autores encuentran que la confianza puede generar la mayoría de la varianza de los resultados académicos que suele ser explicada por la autoeficacia y los demás autoconceptos (Morony y otros, 2013).

Al ser la autoconfianza una suma de respuestas que el alumno se da sobre su capacidad de realizar unas determinadas tareas en las que dividimos el proceso de aprendizaje, los exámenes se pueden convertir en un instrumento para medirla, tal como vimos que hicieron Morony y otros en 2013. Eso es exactamente lo que se pretende con los sistemas de evaluación basados en la confianza, desde el modelo original de Kate Hevner de 1932 hasta el TBC, que estamos estudiando en este trabajo, aunque, como hemos dicho antes, no solo se mide la confianza, sino que esta es un componente de la nota que el estudiante obtiene. En este caso, la confianza lo es en **haber adquirido las competencias cognitivas que el dominio de la materia demanda**. Tras responder a cada pregunta del examen, sea cual sea el formato de éste, el alumno responde a una pregunta sobre su grado de seguridad en haber dado la respuesta correcta, es decir, sobre el grado de confianza de que sus conocimientos sobre la materia concreta que se está evaluando sean los adecuados. Los exámenes con puntuación basada en el grado de confianza serían una encuesta sobre autoconfianza simultánea al examen. Pero como encuesta de confianza, se diferencia en varios puntos de las encuestas habituales, como la mencionada ABC que hemos visto antes:

- En primer lugar, evalúan una confianza retrospectiva, sobre la respuesta ya dada a una pregunta, y no prospectiva. No es la confianza de poder hacer bien algo «en el futuro»,

sino de haberlo hecho bien ya, en ese momento, lo que quiere decir en el pasado más inmediato.

- En segundo lugar, dado que la opción de confianza escogida en esa peculiar encuesta condiciona la nota que el alumno tendrá en el examen, el estudiante se juega su nota al responder a cada pregunta sobre ella; en las encuestas habituales el alumno no se juega nada al responder, salvo su autoconcepto por haber sido o no sincero. Esta repercusión en la puntuación de la respuesta a la encuesta introduce nuevas dimensiones, además de la mera autoconfianza; introduce factores estratégicos y de logro, cálculos sobre las puntuaciones que el estudiante puede obtener. De manera que los exámenes basados en confianza no miden solo dos dimensiones, conocimiento y confianza, miden una tercera dimensión; una dimensión estratégica, de cálculo de puntuaciones posibles, que influye en su elección del nivel de seguridad.
- Como explicaremos más adelante (capítulo 2.3.6.), la sinceridad del alumno en los exámenes basados en la confianza suele ser premiada con un aumento de la puntuación, mientras que la falta de sinceridad suele ser castigada con descenso en la nota. El sistema de examen basado en la confianza es una encuesta que prima la sinceridad más que ningún otro tipo de encuesta concebible.

Partiendo de los condicionantes que acabamos de exponer, los exámenes con puntuación basada en la confianza, habitualmente pruebas de respuesta múltiple (test), suponen el siguiente proceso: Junto con la prueba objetiva, el alumno responde a una encuesta que globalmente le haría la siguiente pregunta: «¿Confías en que tu nivel de conocimientos sobre la materia que estamos evaluando sea el adecuado?». Esa encuesta tiene tantos ítems como preguntas tiene el examen y en cada ítem se evalúa la confianza en los conocimientos del alumno sobre la parcela concreta que se cuestiona. El conocimiento sobre la materia está parcelado en 40, 60, 100 parcelas... etc., según el

número de preguntas del test, y sobre cada una de esas parcelas se hace la misma pregunta, «¿estás seguro de que te lo sabes?». La suma de las puntuaciones sobre la confianza en cada pregunta da la medida de la confianza global del estudiante en su dominio cognitivo de la materia, de la misma manera que el conjunto de preguntas acertadas y falladas da la medida de sus conocimientos reales. **Si la confianza en su dominio de la materia se corresponde con su dominio real de la materia, el alumno sale beneficiado en su nota**, recibe una prima de puntuación. Si la confianza del alumno no se corresponde con sus conocimientos reales, por exceso, «sobreconfianza», o defecto, «infraconfianza», es penalizado por ello.

El sistema de exámenes basados en la confianza transforma una prueba que evalúa una sola dimensión, conocimientos, en una prueba multidimensional, al menos bidimensional -conocimiento más automonitorización del conocimiento-, obliga al alumno a una exigente autorreflexión –exigente porque de sus resultados depende su nota en el examen-, lo que enriquece la evaluación. Este enriquecimiento evaluativo es especialmente significativo al ser, como dijimos antes, la confianza una variable no cognitiva muy robusta, consistente y relacionada con el rendimiento académico y el aprendizaje. La confianza es la medida más ajustada de la automonitorización cognitiva del alumno (Morony y otros, 2013) y, en consecuencia, un excelente medidor de los procesos metacognitivos. Añadiendo la confianza a la medición del conocimiento o de las competencias obtendremos, previsiblemente, un sistema de evaluación superior al tradicional basado en la medida solamente estos últimos. Más adelante hablaremos de la relación que esto tiene con los enfoques y estrategias de aprendizaje de los estudiantes.

2.3. ENFOQUES DE APRENDIZAJE Y EVALUACIÓN

2.3.1. ENFOQUES DE APRENDIZAJE

Desde los trabajos de Marton y Säljö (1976, a y b), se viene utilizando el término «enfoques de aprendizaje» para referirse a la adaptación de estrategias de estudio que llevan a cabo los estudiantes para afrontar sus actividades (Recio, 2005). Los enfoques de aprendizaje son la representación de la forma en que los estudiantes encaran sus actividades académicas, tanto en sus motivaciones como en sus estrategias concretas (Torre, 2007). Este concepto inicial fue desarrollado por Biggs (1978, 1979), en Australia-Hong Kong, y Entwistle (Entwistle, Hanley y Ratcliffe, 1979), en Inglaterra (grupo de Edimburgo), para elaborar sus modelos de aprendizaje. Estos autores crearon cuestionarios de opinión para analizar las estrategias y enfoques de aprendizaje de los alumnos. A partir del cuestionario elaborado por Entwistle, el estudio factorial de las respuestas detectó tres dimensiones o enfoques (Hernández, 1993; Hernández, García y Maquilón, 2001):

- 1) El enfoque profundo/aprendizaje por comprensión/motivación intrínseca.
- 2) El enfoque superficial/aprendizaje por operación/motivación extrínseca/miedo al fracaso.
- 3) El enfoque estratégico/método de estudio organizado/motivación de logro.

El enfoque profundo hace referencia a la búsqueda por parte del alumno de significados y relaciones, y al análisis racional de los contenidos. El aprendizaje se sustenta en la comprensión amplia de la materia estudiada y la motivación suele ser intrínseca -el deseo de aprender-.

El enfoque superficial trata de memorizar información, detalles que consideran importantes para responder preguntas en lugar de buscar significados globales (Kember y Harper, 1987). Su motivación es extrínseca y no es el deseo de aprender, sino el miedo al fracaso lo que mueve al alumno.

El enfoque de logro se orienta a la utilización de los procedimientos y estrategias más eficaces para superar la materia y conseguir los objetivos académicos.

Biggs, por su parte, desarrolló una amplia investigación sobre la teoría de los enfoques de aprendizaje de los estudiantes. Definió a los enfoques de aprendizaje como «los procesos de aprendizaje que emergen de la percepción que el estudiante tiene de la tarea académica influida por las características del individuo» (Biggs, 1987). Elaboró en 1982 el cuestionario SPQ (Study Process Questionnaire), de valoración de enfoques de aprendizaje, con 42 ítems con tres escalas (enfoque superficial, profundo y enfoque de logro), cada una de ellas dividida en dos subescalas (motivo y estrategia) de 7 ítems cada una (Biggs, 1982). Encontró que en cada enfoque había dos tipos de ítems, los que hacían referencia a los motivos del alumno y los que hacían referencia a las estrategias; así había dos subfactores, *motivo* y *estrategia*, en cada uno de los tres agrupamientos, *profundo*, *superficial* y *de logro* (Biggs, 1987). En función de ello elaboró dos cuestionarios; uno, el SPQ (Study Process Questionnaire) de 42 ítems para estudiantes universitarios, y el otro, LPQ (Learning Process Questionnaire), de 41 ítems, para estudiantes de educación secundaria. Posteriormente, y después de que diversos estudios pusieran en cuestión la consistencia de los seis factores de este cuestionario (Hernández, García y Maquilón, 2001), especialmente el factor *de logro*, o *estratégico*, pues algunos de sus ítems tendían al agrupamiento superficial y otros al profundo-, Biggs eliminó este último. De esta manera, elaboró un nuevo cuestionario con solo dos

factores, *superficial* y *profundo*, divididos en dos agrupamientos (o *subfactores*) cada uno, *motivos* y *estrategia*: el R-SPQ-2F, o *cuestionario SPQ revisado con dos factores* (Biggs, 2001; Torre, 2007, p. 200–201).

2.3.1.1. ENFOQUES SUPERFICIAL Y PROFUNDO

Los dos enfoques básicos en la actitud de aprendizaje del alumno, una vez que tiende a descartarse el enfoque de logro, son: un enfoque destinado a la reproducción de los conceptos, es el enfoque memorístico o *superficial*, y un enfoque que busca la comprensión de lo que se aprende, el enfoque de aprendizaje *profundo*. El enfoque superficial de aprendizaje está basado en la repetición de los conceptos hasta que estos quedan fijados y pueden ser reproducidos en los exámenes; no estimula la *meta-memoria* ni, en consecuencia, el aprendizaje de calidad (Pérez, Valenzuela, Díaz y otros, 2011). El enfoque profundo se esfuerza en establecer enlaces lógicos entre los conocimientos adquiridos, la apropiación personal de estos y la búsqueda de significado (Pérez, Valenzuela, Díaz, González-Pienda y Núñez, 2011; Rosario, 2005 y 2007; Torre, 2005). Dadas las definiciones anteriores de enfoque *profundo* y *superficial*, consideramos al primero el más útil para un aprendizaje de calidad, por lo que tendemos a considerar que uno de los objetivos docentes es que el alumno desarrolle enfoques profundos de aprendizaje, basados en la comprensión y búsqueda de relaciones y significados. En realidad, los enfoques profundos pueden ser buenos predictores de resultados en tareas de evaluación continua, pero como predictores de los resultados en exámenes y presentaciones orales algunos autores encuentran que resultan algo pobres (Recio, 2005; Swanberg, 2010; Vrugt, 2008). Algún estudio muestra que los enfoques

superficiales suelen ser débilmente predictores de bajos resultados académicos (Vrugt, 2008). Otros autores (Ekins 1991; Gargallo, 2006; Papinczak, 2009; Svenson, 1977; Torre, 2005) encuentran correlaciones significativas entre los enfoques profundos o profundo-estratégicos de aprendizaje y unos mejores resultados académicos, aunque la intensidad de las correlaciones no es muy grande. Entwistle y Ramsden encuentran correlación entre los enfoques de aprendizaje y la autopercepción de progreso académico de los alumnos (Entwistle y Ramsden, 1983; Ramsden y Entwistle, 1981), pero ellos mismos reconocen la limitación de su estudio, pues al ser los resultados académicos autopercebidos y no objetivos, podría haber un cierto grado de circularidad en el razonamiento, con la previsible correlación entre autopercepción de progresos académicos y autopercepción de enfoques de aprendizaje. Como estamos viendo, aún es necesario completar la evidencia de que los enfoques profundos se relacionen con un alto rendimiento académico, aunque el sentido común nos diga fuertemente que así ha de ser (Pérez y otros, 2011; Recio, 2005).

Los enfoques de aprendizaje del alumno no son inmutables; alumnos que habitualmente emplean enfoques de aprendizaje profundos pueden adoptar ocasionalmente un enfoque superficial si en una materia concreta el enfoque profundo no es demandado por el profesor o es difícil de conseguir en un medio poco propicio al aprendizaje (Kember, 1987). Influyen notablemente a este respecto la motivación intrínseca, la falta de interés del alumno por la materia (Fransson, 1977), la forma que tiene el profesor de enseñar (Ramsden y otros, 1981), la forma de evaluación -si se realiza mediante preguntas que preconizan el componente memorístico- (Entwistle y otros, 1983, cap. 8), el exceso de trabajo para el alumno en épocas de gran concentración de exámenes o la ausencia de libertad para escoger el modelo de aprendizaje en el ambiente académico (Entwistle y otros, 1983; Kember, Charlesworth,

Davies, McKay y Sivan, 1997). Por otro lado, se han encontrado correlaciones positivas entre enfoques profundos de aprendizaje y la autorregulación (De la Fuente, 2008).

Nosotros creemos que las relaciones entre los sistemas de evaluación y los enfoques de aprendizaje son una importante materia de investigación. Un sistema de evaluación de gran calidad sería aquel que, además de proporcionar información sobre los conocimientos del alumno, **primase en la nota los enfoques profundos de aprendizaje** frente a los superficiales (Tian, 2007). Ese procedimiento de evaluación estaría ponderando no solo los conocimientos sino también la comprensión. Uno de los objetivos del presente trabajo será el de comprobar si eso se da con el test basado en la confianza.

2.3.1.2. ENFOQUES DE APRENDIZAJE Y METACOGNICIÓN

El tema de los enfoques de aprendizaje está estrechamente conectado con un aspecto del aprendizaje al que hemos hecho referencia antes, que es el de la *metacognición*. Como hemos dicho, metacognición es el conocimiento que el sujeto tiene de sus propios campos cognitivos, el *conocimiento del propio conocimiento*. El aprendizaje profundo no se limita a fijar en la memoria una serie de conceptos o imágenes, sino que los comprende e interpreta; es decir, delimita los conceptos aprendidos, lo que conlleva conocer su completitud, sus relaciones con los demás conocimientos, su encuadre en un marco lógico y operativo... todo eso es lo que significa realmente «comprender». Los enfoques de aprendizaje profundos suponen, por lo tanto, un proceso constante de autorreflexión a medida que se adquieren conocimientos, suponen adquirir no solo cognición, sino también metacognición. El

aprendizaje superficial, centrado en la reproducción automatizada de contenidos, no se cuida de reflexionar sobre los conceptos recién adquiridos, simplemente los archiva para reproducirlos en el examen, donde termina toda su función.

Pero el proceso de autorreflexión acerca del aprendizaje puede ser correcto o incorrecto. Puede haber resistencias al cambio de tareas o tácticas, aunque las elegidas estén resultando inadecuadas, o darse un error general en la evaluación final del resultado, lo que a veces lleva al alumno a pensar que dedica la estrategia, el tiempo y esfuerzo adecuados a prepararse una materia y que sus fracasos repetidos se deben la forma de evaluar del profesor. Como dijimos antes, esos errores se deben a un deficiente proceso de automonitorización.

Zimmerman cree que, además de los factores metacognitivos que enfatizan los estados de conocimiento y razonamiento deductivo, otros factores, como las creencias y juicios sobre uno mismo, así como la competencia autopercebida, explicarían en gran medida los esfuerzos y la elección de tareas por parte del estudiante (Zimmerman, 1995).

La confianza del alumno, su seguridad o inseguridad, a la hora de aplicar sus conocimientos, respondiendo a las preguntas de un examen, depende del dominio que tenga de su campo cognitivo, del conocimiento que tenga de los límites de este, su completitud y sus lagunas, de que los conocimientos estén incardinados en un marco racional, lógico y funcional que les dé soporte; dependen, en suma, no solo de los conocimientos adquiridos en sí, sino de la *monitorización metacognitiva*. Los procedimientos de examen que evalúan conocimiento y confianza evalúan, por lo tanto, cognición y metacognición. Siendo la metacognición un elemento vinculado a la

aplicación práctica de los conocimientos, el TBC y los demás procedimientos de evaluación basados en la confianza deberían tener capacidad para valorar no solo la «gnosis», la *cantidad* de conocimiento, sino la «praxis», la capacidad de utilizarlo.

2.4. EXAMEN DE TIPO TEST BASADO EN LA CONFIANZA (TBC)

2.4.1. JUSTIFICACIÓN DE ESTE TIPO DE EVALUACIÓN

Como vimos al principio de este marco teórico, en la pirámide de Bloom, el mero conocimiento es solo la base del proceso de aprendizaje; por encima están la comprensión, aplicación, análisis, síntesis y, en lo más alto, la evaluación (Bloom, 1956). En nuestra opinión, todo buen sistema de evaluación debería cubrir, al menos, los dos procesos de la base: el conocimiento y la comprensión, y además, debería ser útil para la autoevaluación del alumno. Cumpliendo esos tres requisitos, probablemente los demás, específicamente el análisis y la síntesis, quedarían también, de alguna manera, cubiertos. Juzgamos que el sistema habitual de pruebas objetivas de respuestas múltiples que veníamos utilizando desde siempre en nuestro centro educativo, el test con puntuación basada en el número de aciertos y con la corrección clásica de la adivinación por azar, no cumplía bien estas funciones; concretamente, estimulaba solo el conocimiento, con escaso estímulo de la comprensión, del análisis y de la síntesis. Por ello decidimos modificar el sistema de exámenes de tipo test que veníamos utilizando.

Desde hace seis años nosotros estamos utilizando en nuestro centro docente el test con puntuación basada en el grado de seguridad, o, como lo vamos a llamar a partir de

ahora, *test basado en la confianza* (TBC). Utilizamos el TBC porque pensamos que ofrece ventajas sobre el sistema de corrección tradicional; fundamentalmente porque da una mayor información sobre el dominio que tiene el alumno de la materia, porque le estimula a estudiar más y porque tiene un gran potencial como procedimiento de autoevaluación para el propio estudiante. Al ser un sistema de examen poco conocido y utilizado en nuestro país vamos a explicar su sistemática y las distintas modalidades que del TBC hay.

2.4.2. DESCRIPCIÓN GENERAL DEL PROCEDIMIENTO

En el *test basado en el grado de seguridad*, o *test basado en la confianza* –TBC–, se elaboran las preguntas habituales del test, con dos o más respuestas, y tras la elección del alumno en cada ítem, se le interroga sobre su grado de seguridad de haberla respondido correctamente (ver nuestro modelo en el anexo D). La puntuación que el alumno obtendrá por cada pregunta correcta dependerá del grado de confianza en acertar que haya expresado. Será máxima con máxima confianza, menor cuando la seguridad expresada en la pregunta sea menor. Cuando la respuesta del alumno resulte ser incorrecta, la penalización que recibirá por ello irá también en función del grado de confianza que haya declarado. Será una penalización alta si el alumno declaró estar seguro de acertar (con frecuencia se le descontará tanto o más de lo que se le hubiera sumado si la hubiese respondido correctamente); será muy baja o nula si el alumno, en cambio, declaró no estar seguro de la respuesta (incluso en uno de los modelos que se utilizan de TBC, como luego veremos, cuando el alumno falla la pregunta, y manifiesta su inseguridad, se le suma una pequeña puntuación).

Con el TBC, el alumno ha de ponderar su seguridad en la respuesta elegida. Es decir, si responde con mayor seguridad recibirá una puntuación mayor, pero se arriesgará a un mayor descuento si yerra; en cambio, si declara menor seguridad, obtendrá menos puntuación si acierta, pero también menor (o incluso nula) penalización si falla. Es, a la vez, un examen y una encuesta de confianza, pero con la peculiar característica de que ambos, respuesta a las preguntas y confianza, conforman la nota que se obtendrá.

2.4.3. OBJETIVOS Y VENTAJAS DEL PROCEDIMIENTO DE EVALUACIÓN BASADO EN LA CONFIANZA

Como dice Gardner-Medwin (2011), en nuestra sociedad, que resalta por encima de todo la evidencia para resolver las incertidumbres, y muy especialmente lo hace en los campos científico y legal, la existencia de las incertidumbres, de la duda, es obviada, mal entendida e ignorada en el proceso de aprendizaje. La educación tradicional resalta el conocimiento, el conocimiento cierto y seguro, y obvia el *reconocimiento de las incertidumbres*, y el manejo por el alumno de estas. Pero en el mundo real, las incertidumbres, la debilidad de las evidencias, las dudas, son más frecuentes que las certezas. En nuestra sociedad, saturada de información, el reconocimiento de las incertidumbres y de la ausencia de evidencias es, con frecuencia, más importante que la posesión del conocimiento en sí (Gardner-Medwin 2011). Y cuando reconocemos la incertidumbre, es necesario ponderarla, cuantificarla, incluso de forma probabilística, para poder manejarla intelectualmente; y como el conocimiento suele tener un porcentaje de incertidumbre, no hay un adecuado conocimiento sin la destreza necesaria

para reconocer y manejar las dudas. El *indicador de confianza* añadido a las preguntas del examen TBC, intenta que el alumno aprenda a reconocer y manejar sus dudas.

Según el modelo de comportamiento basado en el aprendizaje (*Learning-Behaviour Model*) obtenemos el siguiente cuadrante (tabla 2.5), en el que se relaciona la cantidad calidad y seguridad de la información del sujeto con las acciones subsiguientes.

TABLA 2.5.
Modelo de comportamiento de aprendizaje

Mala información <i>Acción equivocada</i>	Maestría (información correcta y autoconfianza) <i>Acción adecuada</i>
Desinformación <i>Parálisis</i>	Conocimiento inseguro <i>Vacilaciones</i>

Se puede apreciar que por debajo de la falta de información está, como situación más perjudicial, la de la información errónea; si la primera induce parálisis en la acción, que se puede resolver buscando ayuda, la segunda induce acciones equivocadas, lo que es bastante más nocivo (Hunt 2003). Por otro lado, tener la información pero carecer de la seguridad induce una acción vacilante e insegura, a diferencia de la auténtica maestría.

Estar confiado en la certeza de los propios conceptos conlleva no solo saber, sino también *saber hacer* cuando del conocimiento se derive alguna actividad práctica. La inseguridad, las dudas sobre el propio conocimiento llevan a la parálisis en la acción. En el campo sanitario esta afirmación resulta muy evidente: si el sanitario no está totalmente seguro de que la conducta a adoptar es la correcta, y por lo tanto, no sabe si beneficiará o perjudicará al paciente, es probable que no actúe en ningún sentido; en la duda, si sabe reconocer su propia incertidumbre se sentirá inducido a buscar más información o a pedir ayuda. No es extraño, por lo tanto, que los exámenes basados en la confianza se utilicen abundantemente en la enseñanza de las Ciencias de la Salud en países de habla inglesa (Barr y Burke, 2013).

Una primera razón para introducir el sistema de exámenes con puntuación basada en la confianza es la de ayudar a los estudiantes a reflexionar sobre la firmeza del conocimiento adquirido; ello hace referencia a su nivel de comprensión profunda de la materia. Un alumno está completamente seguro de la respuesta elegida cuando es capaz de razonarla, de dar argumentos científicos o lógicos a favor de ella, y mucho más si también puede razonar por qué las otras respuestas son falsas; de manera que la puntuación en este tipo de exámenes tiene, de alguna forma, relación con el aprendizaje profundo, que hace hincapié en significados y relaciones de los conocimientos, más que en la memorización de datos.

Otro motivo importante para implantar el TBC en las Ciencias de la Salud es el deseo de que los estudiantes adquieran conciencia sobre el riesgo que supone manejar conceptos erróneos, especialmente si estos están afianzados «con total seguridad». En el campo sanitario evitar conceptos falsos arraigados puede ser de importancia vital: podrían motivar una mala práctica profesional en el futuro, con riesgos para los

pacientes... además, los errores de concepto adquiridos en las etapas formativas pueden constituirse en un gran obstáculo para la adquisición de mayores niveles de conocimiento. Es por ello que un sistema de evaluación que penaliza con mayor intensidad los conceptos erróneos afianzados, ofrece ventajas adicionales sobre los sistemas habituales de test, donde tienen la misma penalización el error por desconocimiento que el error por conocimiento equivocado (Rippey y Voytovich, 1983; Gardner-Medwin, 2006).

Una razón adicional para preferir este tipo de exámenes es la de dotar al profesor de un instrumento de medida más completo y potente para valorar el proceso de aprendizaje del estudiante que el test convencional, valorado habitualmente solo según el número de aciertos, que tiende a medir únicamente el mero conocimiento (Hunt, 2003). Al incluir múltiples posibilidades de puntuación por cada pregunta, acertada o fallada, el sistema basado en la seguridad proporciona una excelente medida del aprendizaje cognitivo (King y Witt, 2009) y puede dar una idea mejor del proceso de aprendizaje profundo que el sistema convencional de test.

2.4.4. EL TBC COMO PROCEDIMIENTO PARA DISTINGUIR EL CONOCIMIENTO SEGURO DEL INSEGURO

Según Morales (2006a) a la hora de responder a una pregunta en un test, el alumno puede estar en alguna de las siguientes situaciones, que exponemos en la tabla 2.6.

TABLA 2.6.

Situaciones que se presentan al responder un alumno a una pregunta de test.

<i>Responder</i>	<i>Con plena seguridad</i>	<i>Con duda</i>	<i>Por casualidad</i>
correctamente	Realmente <i>saben</i> y <i>entienden</i>	Conocimiento <i>correcto pero inseguro</i>	<i>No saben</i> , pero han tenido suerte
incorrectamente	Conocimiento <i>incorrecto</i>	Conocimiento <i>incorrecto pero inseguro</i>	<i>No saben</i> y además no han tenido suerte

Nota. Morales, 2006a.

Como dijimos más arriba, *conocimiento seguro* es aquel en donde, además de recordar el dato concreto por el que se pregunta, el alumno entiende por qué esa es la respuesta correcta y sería capaz de dar razones de ello. En el proceso de responder a una pregunta del examen, de forma correcta o incorrecta, se van a dar las cuatro situaciones siguientes, que exponemos en la tabla 2.7.

TABLA 2.7.

Cuadrante de posibilidades al responder a una pregunta.

	CON SEGURIDAD	SIN SEGURIDAD
Respuesta correcta	Ciencia	Incertidumbre
Respuesta incorrecta	Autoengaño	Ignorancia

Nota. Morales, 2006a.

En este esquema es evidente que la situación de ignorancia no es la más desfavorable, sino la de autoengaño, la de tener un concepto erróneo afianzado con total seguridad.

Con este sistema, la puntuación de cada pregunta acertada pasa de tener un único valor positivo a tener varios valores (tres en nuestro caso), de acuerdo con el dominio que el alumno tenga de la materia. Igualmente se establece una gradación de la penalización por pregunta fallada, penalizando especialmente el exceso de confianza en conceptos equivocados o inseguros. Esto le proporciona al profesor una información

sobre el conocimiento y comprensión de la materia por el alumno mayor que el sistema convencional, con una sola puntuación por cada pregunta (Ebel, 1968).

El reflexionar sobre su saber lleva al alumno a buscar relaciones con otros conceptos; nosotros creemos que este tipo de evaluaciones puede premiar al alumno con enfoques profundos de aprendizaje frente al que tiene enfoques superficiales, propiedad que habría que exigirles a todos los buenos sistemas de evaluación (Tian, 2007). Si se acostumbra a este sistema de evaluación es de esperar que el alumno estudie pensando más, profundizando más y relacionando conceptos.

Con el sistema de test basado en la confianza se espera que el alumno adquiera mejores estrategias de estudio y una mayor conciencia de sus lagunas formativas. Ante cada pregunta del test ya no importa sólo *acertar o fallar*, sino *saber, estar inseguro o no saber*. Los alumnos aprenden a reflexionar sobre su saber, a preguntarse el porqué de los conceptos aprendidos para afianzar su seguridad en la respuesta: no basta un recuerdo inmediato desvinculado. A esta capacidad de juzgar el propio saber Kleitman y Stankov (2001), la denominaron «automonitorización», y la definieron como la habilidad para aprehender y juzgar la calidad del propio trabajo cognitivo. La medida de la confianza del estudiante es una medida ajustada y precisa de la «automonitorización» (Sharma 2011), por lo que incluirla en el proceso evaluativo nos parece muy pertinente. Ya expusimos anteriormente la importancia capital de la automonitorización en el aprendizaje autorregulado (Panadero y otros, 2014).

2.4.5. IMPORTANCIA QUE EN LA EVALUACIÓN TIENE LA CONFIANZA DEL ESTUDIANTE

La *confianza* es una medida relativamente nueva dentro del campo del autoconcepto. La confianza en el propio conocimiento indica un nivel superior de dominio de una materia y, probablemente, una mayor capacidad para aplicar esos conocimientos en la práctica.

Se ha comprobado que la confianza es una dimensión robusta que explica diferencias individuales en el rendimiento académico. Como expusimos antes, en estudiantes de secundaria es el mejor predictor de resultados en matemáticas e inglés (Morony y otros, 2013; Stankov y otros, 2012), física (Sharma y Bewes, 2011) y ciencias (Chang y Cheng, 2008) frente a otros parámetros como autoeficacia docente, ansiedad o autoconcepto. De hecho, está estrechamente relacionada con la autoeficacia y el autoconcepto y es, en realidad, responsable de una gran parte de la varianza predictiva de estas variables, que hasta ahora eran consideradas los mejores predictores de resultados académicos (Morony y otros, 2013; Stankov y otros, 2012).

La confianza es, en realidad, una precisa y ajustada medida de la «automonitorización» del alumno; la mejor forma directa de medirla (Sharma y otros 2011). La comparación entre la confianza expresada por el alumno en la exactitud y corrección de sus conocimientos y la real corrección y exactitud de estos, tal como se pone de manifiesto en el examen, proporciona la *calibración* de su «automonitorización» (Kleitman y otros, 2001; Pallier, 2003). La capacidad de monitorizar efectivamente los procesos cognitivos le permite al estudiante mejorar las estrategias de aprendizaje y le proporciona también una capacidad superior para

reconocer y rectificar las deficiencias propias en conocimientos (Karabenik, 1996); le permite una ajustada autorregulación.

En el test basado en la confianza el alumno podrá maximizar su nota *solamente si su conocimiento real es similar a su confianza en su propio conocimiento* (Rippey, 1978). Dicho de otra manera, si su metacognición es la adecuada. A esta diferencia entre conocimiento y confianza algunos autores la llaman «sesgo de de puntuación» (Moore y Healy, 2008; Stankov, 2000). El fallo en esa parte reguladora del conocimiento, que es la metacognición, llevaría al alumno a cometer errores a la hora de escoger su nivel de seguridad; no es suficiente con saber escoger la respuesta correcta, hay que saber hasta qué punto se conoce la materia evaluada, la densidad y los límites del propio conocimiento.

El sistema de evaluación basado en el grado de seguridad hace posible para el examinador separar la evidencia del conocimiento de la confianza en el conocimiento. Esta separación tiene el potencial de poder proveer mediciones más fiables y válidas del aprendizaje, así como información de la tendencia a sobrevalorar o infraevaluar el conocimiento de cada uno de los alumnos (Rippey, 1978). Algunos autores, como Ahlgren (1969), han encontrado evidencias de que el conocimiento que es señalado por el alumno como “seguro” en este tipo de test permanece más tiempo que el inseguro, con lo que los procedimientos de puntuación basados en la seguridad darían también información sobre el aprendizaje a largo plazo de alumno. El test basado en el grado de seguridad proporciona al docente un método de evaluación más potente que el sistema habitual de test basado solo en el número de aciertos.

Por otro lado, al permitir el TBC al alumno obtener puntuación en preguntas donde tiene conocimientos pero está inseguro, y que con otro sistema de evaluación podría escoger dejar en blanco, permite evaluar el conocimiento parcial, insuficientemente ponderado en los sistemas tradicionales de examen de tipo test.

Además de las ventajas para el profesor, este procedimiento de evaluación le da al alumno la ventaja de poder graduar su riesgo, en función de sus objetivos (asegurar el aprobado o tener una nota alta, por ejemplo) y de su seguridad: no tiene que jugárselo en una pregunta al todo o nada, tiene otras opciones intermedias. El alumno se sentirá así más dueño de su destino en el examen, más *autónomo* en su proceso de aprendizaje.

2.4.6. EL FACTOR DE LA SINCERIDAD DEL ALUMNO AL ESCOGER SU OPCIÓN DE SEGURIDAD

Una ventaja adicional, no desdeñable, del TBC es que este procedimiento de corrección estimula la sinceridad por parte del alumno, penalizando el engaño; así sucederá tanto en el caso de que el alumno escoja un nivel de seguridad superior al real para intentar mejorar su nota (será castigado con mayor penalización en sus respuestas incorrectas) como en el caso de que escoja un nivel de seguridad inferior al real para no asumir riesgos; en este último caso será penalizado también al obtener una menor puntuación en sus respuestas correctas. Esta diferencia entre la confianza y el conocimiento real es lo que algunos autores llaman «sesgo de puntuación» o «realismo» y, en la situación de exceso de confianza frente a los conocimientos reales, «sobreconfianza» (Moore y Healy, 2008; Morony y otros, 2013; Stankov, Lee, y Paek,

2009). El TBC y la retroalimentación de los resultados con él son una ayuda a los estudiantes para identificar cuándo ellos mienten en la escala de seguridad y hacerles reflexionar acerca de ello (Gardner-Medwin, 1995). Como señalan Gardner-Medwin y Gahan (2003), «El análisis de las opiniones de los alumnos sobre este tipo de evaluación muestra que, tras un periodo de entrenamiento del alumno en este tipo de test, son muy pocos los que lo manejan mal. Cuando se utiliza en exámenes se ha comprobado que los alumnos son algo menos confiados que cuando lo utilizan como autoevaluación -que los autores hacen preferentemente online-, esto tiende a reducir algo la nota obtenida por el estudiante, lo que los autores corrigen haciendo ajustes automáticos». La práctica de este tipo de exámenes tiende a mejorarla autoconfianza del estudiante. Un factor a destacar es que los autores que han estudiado el TBC no encuentran diferencias en cuanto a la destreza en el uso de este examen debidas al género (Gardner-Medwin y otros, 2003); otras medidas de la confianza académica, en cambio, sí mostraban una mayor tendencia a la sobreconfianza en los estudiantes varones frente a las mujeres, en el campo de las matemáticas (Morony y otros, 2013). Con la peculiar forma de medir la confianza que es el TBC, y en el campo predominante de las Ciencias de la Salud, no se aprecia esa diferencia entre géneros -cuestión que nosotros también estudiaremos en nuestro trabajo-.

2.4.6.1. VENTAJAS DE UN SISTEMA DE EXAMEN QUE VALORE, ADEMÁS DE LOS CONOCIMIENTOS, LA CONFIANZA

Consideramos que el test basado en el grado de seguridad puede constituir no solo un buen procedimiento de examen sino una útil ayuda al alumno en su autoevaluación

del aprendizaje, en su *autorregulación*. Como sistema para una autoevaluación del alumno a lo largo del curso, antes del periodo de exámenes, proporciona una información más rica al estudiante que un test convencional, pues al preguntarle por su seguridad fuerza al alumno a una autorreflexión más intensa, y le podría estimular a adoptar estrategias de estudio que enfatizan no solo la memorización sino también la comprensión, dado que, como hemos dicho varias veces, uno solo está totalmente seguro de aquello que comprende. No ha sido suficientemente estudiada, hasta ahora, esta faceta del examen basado en el grado de seguridad, la de hasta qué punto este tipo de evaluación puede inducir un cambio de estrategias de aprendizaje, y hacerlo constituye uno de los objetivos del presente trabajo.

La cuestión del grado de seguridad con que el alumno responde a las preguntas de un examen sobrepasa la cuestión de su mera utilidad para la evaluación o la autoevaluación y nos introduce en un aspecto muy relevante del aprendizaje, intensamente relacionado con la confianza del alumno, que es el de la «*monitorización metacognitiva*»; es decir el conocimiento y control del propio conocimiento, de los que depende la seguridad o inseguridad del estudiante en el examen y en la actividad práctica. El TBC nos permite adentrarnos en este campo, la metacognición, en mayor medida que los sistemas tradicionales de evaluación.

2.4.7. MODELOS DE PUNTUACIÓN DE TEST BASADOS EN LA CONFIANZA

2.4.7.1. UN POCO DE HISTORIA

El primer caso publicado de un examen con puntuación basada en el grado de seguridad lo fue por Kate Hevner, profesora de Apreciación Musical en enseñanza

secundaria en Estados Unidos, en 1932. En ese caso se trataba de un examen de test de tipo «verdadero/falso». El procedimiento no vuelve a utilizarse hasta más de 30 años después (Ebel, 1968; Gardner, 1969; Heywood, 1977; Khan, Davies y Gupta, 2001). En tiempos más recientes, ha sido el grupo de Gardner-Medwin (1995, 1998, 2006, 2007), en la escuela de medicina del University College of London, quien ha estudiado más profundamente este sistema. Nosotros vamos a seguir, parcialmente, su método, que en el Reino Unido es muy utilizado en carreras relacionadas con las Ciencias de la Salud (Gardner-Medwin y Gahan, 2003, Barr y otros, 2013).

2.4.7.2. TABLA EXPLICATIVA DE LAS CARACTERÍSTICAS DE LOS SISTEMAS DE TBC

Vamos a exponer los principales sistemas de puntuación de pruebas objetivas basados en la confianza del alumno que se han utilizado por otros autores hasta ahora. Lo vamos a hacer en tablas en las que pondremos las opciones de seguridad que se le ofrecen al alumno y las puntuaciones, positivas y negativas, que se les asignan según la respuesta resulte correcta o incorrecta. Todos los sistemas de test basados en la confianza dan mayor puntuación por respuesta correcta cuanto mayor sea el nivel de seguridad del alumno y restan mayor puntuación por respuesta incorrecta por el mismo motivo; esa es la esencia de su funcionamiento.

El modelo de tabla que utilizaremos para analizar los distintos sistemas es el siguiente (tabla 2.8):

TABLA 2.8.

Modelo de tabla utilizada para analizar los sistemas de puntuación

Niveles de seguridad	alto	medio	bajo
Puntuación que suma la respuesta correcta	+ X	+ Y	+ Z
Puntuación que resta la respuesta errónea	- L	- M	- N
Probabilidad de acertar que hace rentable la opción	100 % - A %	A % - B %	< B %

En las casillas correspondientes a las columnas de *alto*, *medio* o *bajo* ponemos, en la primera fila, los puntos que se suman en cada opción de seguridad cuando el alumno responde correctamente (+X, +Y, +Z en el ejemplo) y en la segunda los que se restan en cada opción cuando la respuesta es incorrecta (- L, - M, - N).

En la última fila ponemos un dato importante para valorar la **coherencia** de estos sistemas: *El intervalo de probabilidad de acertar que haría rentable* (en términos de puntuación) *cada opción de seguridad*; Si en un modelo de test una opción de seguridad no resulta rentable sea cual sea la probabilidad de acertar, lo señalamos poniendo “*nunca*” en su casilla de la última fila; es en este último caso cuando decimos que ese sistema de puntuación es “*incoherente*”. Este es un detalle que la mayoría de los autores no tiene en cuenta y, como veremos en las diferentes tablas de puntuación, la mitad de los sistemas de puntuación del TBC que se han propuesto resultan «incoherentes». Un sistema de puntuación incoherente puede, a pesar de todo, funcionar bien como sistema de evaluación, pero consideramos que la coherencia en las puntuaciones debería ser un requisito de toda buena tabla de puntuación de un TBC.

2.4.7.3. MODELO DE AHLGREN

Ahlgren propuso en 1970 un original sistema de puntuación que combina la puntuación basada en la confianza con el sistema tradicional, con o sin penalización por

respuesta errónea. En este sistema hay tres niveles de seguridad, *seguro*, si el alumno juzga que la posibilidad de acertar es claramente mayor de un 50 %; *neutral*, si la posibilidad de acertar está en torno a un 50 %; *adivinación* (“guess”) si el alumno cree que tiene menos de un 50 % de probabilidades de acertar. Como se ve inmediatamente, este sistema no está pensado para preguntas con dos opciones de respuesta (verdadero/falso), pues en ese caso la probabilidad de acertar es siempre igual o superior al 50 %, sino para test de respuestas múltiples, de tres opciones o más. El sistema consiste en que cuando el alumno escoge la opción de máxima seguridad se le suma o resta 1/3 de la nota correspondiente a la pregunta, según acierte o falle; cuando escoge la opción de menor seguridad se le descuenta 1/3 de la nota a sumar o a restar, según su acierto o fallo, y cuando escoge la opción *neutral* recibe la puntuación prevista para esa pregunta sin adiciones ni sustracciones. La tabla de puntuación con el sistema de Ahlgren sería la siguiente:

TABLA 2.9.
Modelo de TBC de Ahlgren (1970)

	<i>Seguro</i> , probabilidad de acertar > 50 %	<i>Neutral</i> , probabilidad de acertar \approx 50 %	<i>Adivinación</i> , probabilidad de acertar < 50 %
correcta	sumar 1/3	igual	restar 1/3
incorrecta	restar 1/3	igual	sumar 1/3

Nota. La suma o la resta, se hace de la puntuación asignada a esa pregunta.

2.4.7.4. MODELO DE HASSEMEN Y HUNT

Otro modelo de puntuación en un test basado en la confianza es el propugnado por Hassmen y Hunt en 1994 y por Khan (2001). Para preguntas con dos opciones de respuesta, el sistema considera 5 niveles de seguridad, con puntuaciones positivas para

las respuestas correctas que van desde un 20 % para el nivel mínimo hasta el 100 % en el máximo, y para las respuestas incorrectas oscilan entre el -120 % para el nivel máximo y el +10 % para el mínimo. Aparte de la incoherencia que consideramos que hay en dar una puntuación positiva para una respuesta errónea, se considera que cinco niveles de seguridad son demasiados para que sean intuitivamente comprendidos por la mayoría de los alumnos (Gardner-Medwin, 2006), mucho más con las complejas cifras de puntuación que hay en este sistema. La tabla 2.10 muestra los niveles de seguridad y puntuaciones del sistema de Hassmen y Hunt para preguntas con dos respuestas (verdadero/falso).

TABLA 2.10.

Modelo de TBC de Hassmen y Hunt (1994) y Khan (2001)

Niveles de seguridad	5	4	3	2	1
Respuesta correcta	+ 1	+ 0.9	+ 0.74	+ 0.54	+ 0.2
incorrecta	- 1.2	- 0.64	- 0.32	- 0.08	+ 0.1
Intervalo rentable	100 %– 84'8%	84'8 %– 66'7 %	66'7 %– 54'5 %	54'5 %– 50%	Nunca*

Nota. *Con dos opciones, la probabilidad mínima de acertar es de un 50 %

Es de destacar la característica de que, a pesar de que este procedimiento da una puntuación positiva (+0.1) a las preguntas falladas con el nivel mínimo de seguridad, nunca resultaría rentable para el alumno escoger esta opción de mínima seguridad.

2.4.7.5. MODELOS DE DAVIES

Davies en 2002 propone, en función del número de respuestas por ítem, tres modelos de puntuación para este tipo de exámenes. El modelo «A» correspondería a un test verdadero/falso, el «B» a un test con tres respuestas por ítem y el «C» a uno con 4 respuestas por ítem. Los exponemos en las tablas 2.11, 2.12 y 2.13.

TABLA 2.11.

Modelo de TBC, A de Davies (2 opciones de respuesta por ítem)

Niveles de confianza:	Alto	Medio	Bajo
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta incorrecta	-3	-2	-1
Porcentaje de aciertos en que resulta rentable	100 %– 50 %	<i>nunca</i>	<i>nunca</i>

Nota. (Davies, 2002)

TABLA 2.12.

Modelo de TBC, B de Davies (3 opciones de respuesta por ítem)

Niveles de confianza:	Alto	Medio	Bajo
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta incorrecta	-2	-1	0
Porcentaje de aciertos en que resulta rentable	100 %– 40 %	<i>nunca</i>	< 40 %

Nota. (Davies, 2002)

TABLA 2.13.

Modelo de TBC, C de Davies (4 opciones de respuesta por ítem)

Niveles de confianza:	Alto	Medio	Bajo
Respuesta correcta	+ 5	+ 3	+ 1
Respuesta incorrecta	-2	-1	0
Porcentaje de aciertos en que resulta rentable	100 %– 33.3 %	<i>nunca</i>	< 33.3 %

Nota. (Davies, 2002)

Los tres esquemas de Davies no estimulan al alumno a escoger la opción intermedia, que no resultaría *rentable* en ningún caso; por ello entendemos es un sistema de puntuación incoherente. La opción de máxima seguridad puede ser escogida como opción más racional a pesar de que haya un bajo porcentaje de confianza (50 %, 40 % y 33 % en cada supuesto, respectivamente), lo que a nosotros no nos parece adecuado.

2.4.7.6. MODELO DE SAINSBURY Y WALKER

Sainsbury y Walker (2008) utilizan para el nivel de seguridad una escala de 5 (*seguro al 100 %*) a 1 (*totalmente inseguro*) para preguntas con dos opciones de respuesta con esta clave de corrección. Su modelo figura en la tabla 2.14.

TABLA 2.14.

Modelo de TBC de Sainsbury y Walker, propuesta 1.

Niveles de seguridad	5	4	3	2	1
Respuesta correcta	+ 5	+ 4	+ 3	+ 2	+ 1
Respuesta incorrecta	-2.5	-2	-1.5	-1	-0.5
Porcentaje de aciertos en que resulta rentable	100 % - 33 %	<i>nunca</i>	<i>nunca</i>	<i>nunca</i>	<i>nunca</i>

Nota. Sainsbury y Walker, 2008.

El sistema propugnado por Sainsbury y Walker es más complejo que un simple examen. Se hace la evaluación en dos fases: en una primera los alumnos responden a las preguntas (son pruebas muy breves, de 4 a 8 ítems con dos opciones de respuesta) indicando el grado de seguridad entre los niveles propuestos; en la segunda fase los alumnos discuten las propuestas con sus compañeros y vuelven a responder a las mismas preguntas por segunda vez, también con indicación del grado de seguridad, siendo esta segunda vez la que se califica. Según los autores, al discutir los resultados con sus compañeros los alumnos aumentan su nivel de seguridad y son más los que cambian de respuestas incorrectas a correctas que a la inversa.

Si se analiza esa escala de corrección de Sainsbury y Walker, se observará que hay pocos incentivos para que el alumno escoja las opciones de menor confianza. Si ha reflexionado sobre el sistema de puntuación, debería escoger siempre la opción de máxima confianza sea cual sea su nivel de seguridad, aunque ignore totalmente la respuesta.

Los mismos autores (Sainsbury y Walker, 2008) han hecho otra propuesta de puntuación, en esta ocasión con menos opciones de seguridad, con tres (tabla 2.15)

TABLA 2.15.

Modelo de TBC de Sainsbury y Walker, propuesta 2

Niveles de seguridad	alto	medio	bajo
Respuesta correcta	+ 1	+ 0.3	+ 0.1
Respuesta incorrecta	-1	-0.3	-0.1
Porcentaje de aciertos en que resulta rentable	100 %– 50 %	<i>nunca</i>	< 50 %

Nota. Sainsbury y Walker, 2008.

Esta propuesta incurre en el mismo error que la anterior, hay opciones de confianza («medio»), en las que nunca es rentable para el alumno escogerlas, y para ítems con dos opciones de respuesta la opción racional es optar siempre por el máximo nivel de seguridad.

2.4.7.7. MODELO DE GARDNER-MEDWIN

Gardner–Medwin es el autor que más extensamente ha estudiado los sistemas de examen basados en el grado de seguridad (Gardner-Medwin 1995, 1998, 2005, 2006 a y b, 2007, 2011 ; Gardner-Medwin y Gahan, 2003). El autor propone un sistema de test con tres niveles de *confianza*, o *seguridad*: alto, medio y bajo. No considera que sea práctico poner más de 3 niveles de seguridad, como pasaba con los sistemas de Hassmen y Hunt, pues la excesiva complejidad que conlleva hacer frente a muchas opciones dificultaría la comprensión intuitiva por el alumno; nosotros estamos de acuerdo con esa opinión. Plantea dos sistemas de puntuación; el primero para preguntas con dos opciones de respuesta (verdadero/falso) y el segundo para cuando las opciones de respuesta sean tres o más. El autor suele preferir el test con dos opciones de respuesta

(Gardner-Medwin 1995), pero elaboró también una escala de puntuación para cuando las opciones de respuesta fuesen más de dos. La diferencia entre la puntuación para el test con dos respuestas por pregunta y la puntuación del que tiene más de dos respuestas consiste en la penalización que se aplica a las respuestas incorrectas en los niveles elevado e intermedio de seguridad (-6 y -2 respectivamente en exámenes verdadero/falso, -4 y -1 cuando hay tres o más opciones de respuesta). En cualquier caso, responder incorrectamente con total seguridad está «*sobrepenalizado*», es decir, tiene más penalización que los puntos que sumaría si la respuesta fuese correcta.

Una característica a la que Gardner-Medwin da importancia (Gardner-Medwin y Gahan, 2003) es la de que la relación entre la puntuación positiva por acierto y la negativa por fallo en cada nivel de confianza ha de ser *coherente*, concepto al que hemos hecho referencia antes. Llamamos *coherencia* al hecho de que haya un intervalo en el grado de seguridad del alumno (concebido como el porcentaje de respuestas correctas) en el cual cada nivel de seguridad sea el más favorable. Así, por ejemplo, en el sistema de Gardner-Medwin para test verdadero/falso, cuando la probabilidad de acertar es menor de 67 % la opción de menor seguridad es la más favorable, cuando oscila entre el 67 % y el 80 % la mejor opción es la de seguridad intermedia y cuando la probabilidad de acertar es mayor del 80 % entonces la opción de mayor confianza es la preferible. Cada intervalo de probabilidad de acertar –o riesgo de fallar- tiene su propio nivel de seguridad más adecuado.

En la tabla 2.16 ponemos los modelos de Gardner-Medwin: el diseñado para dos opciones de respuesta (verdadero/falso) y el diseñado para tres o más opciones.

TABLA 2.16.

Modelo de TBC de Gardner–Medwin para 2 opciones de respuesta

Niveles de seguridad	Alto	Medio	Bajo
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta errónea	-6	-2	0
Intervalo en que la opción de seguridad es rentable	100 % - 80 %	80 % - 67 %	< 67 %

Nota. La probabilidad mínima de acertar con 2 opciones es de un 50 % (Gardner-Medwin, 2006).

En este sistema, para dos respuestas, al alumno le resultaría rentable escoger el nivel máximo de seguridad si su posibilidad de acertar fuese superior a un 80 %; si tuviera entre un 67 % y un 80 % de probabilidad de acertar le resultaría preferible escoger la opción de seguridad intermedia, y sólo debería escoger la opción de menor seguridad si su probabilidad de acertar estuviera por debajo del 67 % (con dos respuestas, la mínima probabilidad de acertar es de un 50 %). No necesitamos señalar que el alumno difícilmente puede establecer con tan extremada exactitud su probabilidad de acertar o fallar, pero es una cifra orientativa para él.

TABLA 2.17.

Modelo de TBC de Gardner–Medwin para 3 o más opciones de respuesta

Niveles de seguridad	Alto	Medio	Bajo
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta errónea	-4	-1	0
Intervalo en que la opción de seguridad es rentable	100 % - 75 %	75 % - 50 %	< 50 %

Nota. La probabilidad mínima de acertar con 3 opciones es de un 33.33 % (Gardner-Medwin, 2006).

Con esta tabla de puntuaciones, el alumno debería escoger la opción de máxima seguridad si estima que su probabilidad de acertar es superior a un 75 %, debería escoger la opción intermedia cuando estimara una probabilidad de acertar entre un 50 % y un 75 % y sólo debería escoger la opción de menor seguridad si su probabilidad de acertar estuviera por debajo de un 50 % (lo mínimo, con tres respuestas, es un 33 %).

El modelo de Gardner-Medwin ha sido el modelo seguido por nosotros, pues nos parece fácil de comprender por el alumno y con puntuaciones muy coherentes. Dado que nuestros exámenes de tipo test son siempre de tres opciones de respuesta, nosotros empezamos a aplicar el sistema correspondiente de Gardner-Medwin, con penalizaciones de -4 , -1 y 0 para las respuestas incorrectas. Para orientar al alumno sobre lo que significa cada uno de los niveles de seguridad, es importante hacerle ver que se está *totalmente seguro* de una respuesta cuando, además de creer que es la correcta, se es capaz de explicar *por qué* es esa la correcta y no las otras respuestas; solo cuando el alumno se siente capaz de explicar la respuesta escogida es cuando está *totalmente seguro*. En otro caso, debería escoger las otras opciones de menor seguridad.

2.4.7.8. MODIFICACIÓN DEL SISTEMA DE GARDNER-MEDWIN

Nosotros inicialmente seguimos el sistema de Gardner-Medwin para tres opciones de respuesta, y sobre él desarrollamos el estudio de validez que presentamos como estudio preliminar del presente trabajo. Pero al hacer el análisis de los niveles de seguridad escogidos por los alumnos pudimos constatar que en nuestros alumnos la penalización aplicada a las respuestas incorrectas con máxima seguridad, superior en un 33 % a la puntuación positiva que se obtendría si hubiesen sido correctas, inducía una excesiva prudencia a la hora de escoger el nivel de seguridad, lo que tendía a penalizar las notas, especialmente las de aquellos alumnos con buen nivel de conocimientos. Las entrevistas personales nos confirmaron esta impresión: la «*sobrepenalización*» inducía una prudencia excesiva, perjudicial para la puntuación en los alumnos, especialmente de aquellos que obtenían notas en el intervalo entre 6 y 8 sobre 10; eran alumnos que

tenían confianza en aprobar con las respuestas de las que estaban seguros y que no aspiraban a conseguir una nota de excelencia («matrícula de honor»), lo que les hubiese estimulado a asumir más riesgos. Si bien la prudencia excesiva podría ser considerada un rasgo de la personalidad negativo para el desarrollo pleno de las competencias de un aprendizaje, y, por lo tanto, algo «penalizante», no nos parecía adecuada esa situación, que un rasgo de la personalidad influyera de manera tan clara en la puntuación. La penalización excesiva podría estar, en realidad, acentuando ese rasgo negativo en muchos alumnos. Por todo ello, decidimos rebajar ligeramente la penalización para las respuestas incorrectas con máxima seguridad, de descontar un 133 % de la nota máxima a descontar únicamente un 100 % (-3 puntos en lugar de -4). La diferencia era, de todas maneras, muy pequeña, pero los alumnos se atrevieron a asumir un riesgo algo mayor, manteniéndose, en cualquier caso, un buen efecto disuasorio de la penalización máxima. En la tabla 2.18 ponemos nuestro cuadro de puntuaciones con el TBC.

TABLA 2.18.

Tabla de puntuación utilizada por nosotros desde el año 2012

Niveles de seguridad	Alto	Medio	Bajo
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta errónea	-3	-1	0
Intervalo en que la opción de seguridad es rentable	100 % – 67 %	67 % – 50 %	< 50 %

Nota. Modificada del modelo de Gardner–Medwin para ítems con tres opciones de respuesta (tabla 2.17).

Esta tabla de puntuación -2.18- sigue siendo *coherente* en los tres niveles de confianza. En esta tabla de puntuaciones es algo mayor el intervalo en el que resulta racional escoger la opción de mayor seguridad en comparación con las originales de Gardner–Medwin, como se puede ver. Mientras que en la tabla original era rentable escoger esta opción cuando había un 75 % de probabilidades de acertar, aquí es preferible la opción de máxima seguridad a partir de 2/3 (66.7 %) de probabilidades de

acertar. Por lo mismo, el nivel medio de seguridad tiene un menor intervalo en el que resulta rentable escogerlo; solo es la opción preferible si la probabilidad de acertar oscila entre el 50 % y el 67 %.

2.4.8. VALIDEZ Y FIABILIDAD DEL PROCEDIMIENTO DE CORRECCIÓN BASADO EN LA CONFIANZA

2.4.8.1. INTRODUCCIÓN AL TEMA DE LA VALIDEZ Y FIABILIDAD DEL TBC

Numerosos estudios han mostrado la validez y fiabilidad de estos métodos de examen basados en la confianza, aunque no de forma unánime desde hace más de 40 años; últimamente, en cambio, se han publicado pocos trabajos sobre el tema. Según algunos autores (Gardner, 1969; Hamblen, 1970, Kansup y Hakstian, 1975; Leclercq, 1982), la correlación de los resultados obtenidos en las pruebas basadas en el nivel de confianza con los obtenidos con otros sistemas de evaluación (comprobación habitual de la *validez*) es mejor que cuando se corrige simplemente por el número de aciertos. Para otros (Hevner 1932, Ahlgren 1969, Anderson 1982, Gardner–Medwin 2005), la consistencia interna (o fiabilidad) de los resultados aumenta también. Otros estudios, en cambio, no han mostrado incremento ni en la validez (Rippey, 1972) ni en la fiabilidad de los resultados del examen con este procedimiento (Frary 1982).

En realidad, el sistema clásico de corrección, en el que al alumno se le permite la alternativa de dejar la respuesta en blanco sin sumar puntos pero sin recibir penalización, es un sistema en el que la puntuación también depende en cierta medida de la autoconfianza del sujeto. La diferencia está en que los sistemas de corrección basados en la seguridad dan una puntuación positiva -aunque menor- cuando se acierta

de forma insegura, así se valora el conocimiento parcial y se evita perjudicar a los alumnos más cautos, permitiendo, además, que éstos tomen conciencia de su seguridad y de sus inseguridades.

2.5. RESUMEN SOBRE EL ESTADO DE LA CUESTIÓN. TEST BASADO EN LA CONFIANZA

Las pruebas objetivas con puntuación basada en la confianza (TBC), utilizadas con cierta frecuencia desde hace unos 25 años, son un buen instrumento de evaluación que proporciona ventajas frente a los sistemas tradicionales de exámenes de respuesta múltiple, y que puede resultar útil tanto a los profesores, para conocer mejor el nivel de conocimientos de sus alumnos, como a estos para su autoevaluación. Han demostrado ser un procedimiento de evaluación válido y fiable, un tipo de examen que obliga al alumno a reflexionar sobre la solidez de sus conocimientos, una buena guía de aprendizaje que, incluso, podría orientarle hacia unos hábitos de estudio más profundos (Gardner-Medwin, 2006 a y b; Adams y Ewen, 2009). El TBC puede, además, proporcionar una información que creemos que es crucial para la docencia: el grado de confianza del alumno acerca de sus propios conocimientos. También puede dar información, a los docentes y al propio alumno, sobre su tendencia al riesgo. Se ha podido comprobar en un estudio de cohortes en estudiantes de medicina que el uso del examen basado en la confianza, frente a los exámenes convencionales, dio resultados positivos en términos de mejor rendimiento académico (Schoendorfer y Emmett, 2012). Aún no está totalmente dilucidada su capacidad para detectar los enfoques de aprendizaje del alumno y hace falta un estudio específico en ese sentido, estudio que nosotros hemos emprendido en este trabajo. A pesar de la popularidad de este sistema

de evaluación en los países de habla inglesa, en nuestro país se utiliza todavía muy poco y se echan en falta estudios sobre el tema.

Nosotros hemos adquirido experiencia en la utilización del TBC como procedimiento habitual de examen. Hemos comprobado que genera un cierto sentimiento de rechazo inicial en el alumno que todavía no lo conoce. También hemos constatado que el alumno necesita un cierto entrenamiento en este tipo de test para hacer un adecuado uso de él y no se le puede enfrentar por primera vez con un examen de este tipo en una prueba evaluativa importante sin un proceso de aprendizaje previo. El aprendizaje previo que los alumnos suelen realizar consiste generalmente en la realización de varios test de tipo TBC, habitualmente de muy corta extensión, con retroalimentación de los conocimientos; estos «microexámenes» además sirven de autoevaluación. Otros autores coinciden con nosotros en la necesidad de que el alumno adquiera destreza con este tipo de examen (Gardner-Medwin y Gahan, 2003). Cuando el alumno se acostumbra, el examen basado en la confianza es bastante bien aceptado y, al crearle la sensación de que es una prueba más exigente que el examen de test tradicional, puede motivar un mayor trabajo de estudio en él. Con el TBC se valora no solo conocimientos, sino también confianza; se valora la metacognición, pues discernir el grado de seguridad es, fundamentalmente, un proceso metacognitivo, como dijimos antes. Es también un modelo de examen que se puede adaptar para su corrección mecanizada, aunque esa posibilidad nosotros no la hemos utilizado pues hemos usado siempre la corrección manual. Si se pudiera demostrar que, además de inducir al alumno a dedicar un mayor tiempo al estudio, induce también estrategias de aprendizaje profundas, bien se podría considerar un sistema de evaluación excelente. Comprobar la consistencia interna de estas pruebas, valorar su impacto en la forma de estudiar del

alumno y ver si discrimina los enfoques de aprendizaje profundos y la autorregulación son los objetivos del presente trabajo.

3. METODOLOGÍA DEL ESTUDIO

3.1. OBJETIVO DE LA INVESTIGACIÓN

Un buen sistema de evaluación ha de cumplir varias condiciones. En primer lugar, ha de dar al evaluador una información lo más exacta posible del nivel de conocimientos o competencias del alumno; tiene que ser, para ello, un sistema de probada validez y fiabilidad. En segundo lugar, ha de ser un sistema de evaluación fácilmente comprendido por el alumno y que no le genere un gran rechazo o estrés; un sistema que le provoque temor y le genere excesiva ansiedad no sería deseable. En tercer lugar, ha de dar información al alumno sobre su aprendizaje y orientarle sobre el proceso; es decir, ha de servir para una correcta autoevaluación. En cuarto lugar, debe estimular el trabajo de estudio en el alumno y no dar una expectativa de excesiva facilidad que rebaje el listón de conocimientos a adquirir. En quinto lugar, debe fomentar estrategias y enfoques de aprendizaje profundos con preferencia a los superficiales. En sexto lugar, no ha de suponer una mayor dificultad, o facilidad, a ningún grupo de alumnos en función de circunstancias no académicas; específicamente no debe causar un sesgo de género en la calificación de los alumnos. Nuestro trabajo intenta estudiar si se cumplen esas condiciones en el sistema de test basado en la confianza.

Hicimos inicialmente, antes de comenzar con el cuerpo principal del presente trabajo, un estudio previo sobre la validez y fiabilidad del test basado en la confianza, para comprobar si en nuestro medio educativo se seguían cumpliendo las condiciones de validez y consistencia interna del sistema que habían sido publicadas por autores de otros países. Hay numerosos estudios en otros países sobre el tema, si bien se echan en

falta estudios recientes; los trabajos, en general, han mostrado que el TBC es un sistema válido (Gardner, 1969; Hambelton, 1970; Kansup y otros, 1975; Leclerq, 1982) y fiable (Ahlgren, 1969; Anderson, 1982; Gardner-Medwin, 2005; Hevner, 1932), especialmente en comparación con el sistema tradicional de test, aunque esta opinión no ha sido siempre unánime (Frary, 1982; Rippey, 1972; Zimmerman, 1970), pero no había ningún estudio hecho sobre este sistema de test en España.

Partiendo del estudio preliminar de validez y fiabilidad, que hicimos en 2009, pero que al no estar publicado y ser la base del actual trabajo incluiremos como «estudio preliminar», hemos diseñado el presente trabajo que presentamos, en donde pretendemos profundizar en la línea iniciada. En la investigación, con esa finalidad, nos marcamos los siguientes objetivos:

- Comprobar la validez y consistencia interna del TBC como sistema de evaluación. La validez se estudió en el trabajo piloto realizado en 2009 y también en ese estudio se obtuvieron datos preliminares sobre la fiabilidad, que se completarán en este trabajo.
- Valorar la aceptación del TBC por el estudiante, comparándolo con la aceptación que en él tienen el test convencional y el examen de preguntas abiertas.
- Valorar si el TBC modifica los hábitos de estudio, o *enfoques de estudio*, del alumno, también en comparación con los otros sistemas de evaluación. Se pregunta a los alumnos sobre la modificación que suponen esos tipos de examen tanto en lo referente al esfuerzo o tiempo dedicado a la preparación de cada uno como al tipo de estudio que cada examen genera.
- Analizar la opinión del alumno sobre este tipo de examen, en comparación con los sistemas tradicionales, especialmente en lo referido a lo que constituiría el valor

fundamental de un examen: *la validez percibida como sistema de evaluación* del aprendizaje. Se compara la opinión del alumno sobre el TBC y sobre los test convencionales y exámenes de preguntas abiertas.

- Estudiar los enfoques de aprendizaje de nuestros alumnos, profundos y superficiales según el R-SPQ-2F, y analizar las ~~cor~~relaciones de estos enfoques con los demás ítems de la encuesta.
- Comprobar si el estilo de aprendizaje del alumno, profundo o superficial, supone una modificación en la puntuación obtenida con el TBC en comparación con la puntuación que se obtendría con un test convencional. De esta manera intentaremos evaluar si este examen premia o incentiva un aprendizaje profundo o superficial.
- Comprobar si la autorregulación del alumno influye en la nota obtenida en el TBC más de lo que influye en los test por aciertos.
- Analizar la autoconfianza del alumno, que este tipo de examen permite medir, y ver si se relaciona con los enfoques de aprendizaje, con la autorregulación y con el género de los estudiantes.
- Estudiar la adecuación entre el conocimiento autopercebido por alumno, medido según la PC, y su conocimiento real, medido según el número de aciertos en el test, lo que Gardner-Medwin (2006) llama «sinceridad» del alumno, influyen en la nota obtenida con el TBC.
- Comprobar si el género del alumno influye de alguna manera en la nota obtenida con el TBC en comparación con el test convencional, es decir, si el TBC induce un sesgo de género.

Un sistema de evaluación que fuese válido, fiable, bien aceptado por los alumnos, adecuado para la autoevaluación y estimulante del trabajo del estudiante, de los enfoques de aprendizaje profundos, de la autorregulación y de la sinceridad, y que además no tuviera sesgo de género, podría ser considerado un sistema de evaluación excelente desde todos los puntos de vista. Nos proponemos, por lo tanto, someter a evaluación el examen de tipo test basado en la confianza para ver si cumple esas condiciones. Para poder hacerlo, utilizaremos la metodología que a continuación describimos.

3.2. MUESTRAS UTILIZADAS EN EL ESTUDIO

Los datos para la elaboración de este estudio se recogen de dos fuentes: por un lado de dos series de exámenes hechos por alumnos de enfermería y fisioterapia, y por otro de una encuesta y dos escalas hecha a los mismos. Los exámenes de enfermería y fisioterapia los utilizamos para hacer el estudio de fiabilidad de TBC, lo que constituye el primer objetivo del presente trabajo. La encuesta y las escalas, que son responsables de la mayor parte del presente trabajo, nos permitirán estudiar los otros nueve objetivos.

3.2.1. PARTICIPANTES INCLUIDOS EN EL ESTUDIO DE VALIDEZ

Se han utilizado grupos y participantes diferentes en cada una de las tres partes del estudio empírico. Desde 2008 a 2014 dispusimos de un total de 22 grupos de alumnos – de Enfermería y Fisioterapia- con un total de 1000 estudiantes, aproximadamente. Utilizamos 5 grupos, con 215 participantes, en el estudio de validez; otros 5 grupos, con 254 participantes, en el de consistencia interna, y 4 grupos, con 272 participantes, en la

encuesta y escalas. Los grupos que se utilizaron en cada tipo de estudio eran, con una excepción, independientes de los utilizados en los otros. La utilización de esa diversidad de participantes, en lugar de usar los mismos grupos y participantes para los tres estudios, requiere, tal vez, una explicación.

Ante todo, no hace falta usar el mismo grupo de sujetos para estudiar validez, consistencia interna ni para realizar las encuestas, puesto que son estudios independientes; sí es necesario, en cambio, que el sistema de examen sea el mismo en los tres casos, como efectivamente fue. El uso de grupos diferentes en años distintos fue, por otra parte, inevitable. En el estudio de la validez, que se inició en 2008, se utilizaron los exámenes realizados por los alumnos de tres grupos de Fisioterapia y un grupo de Enfermería en enero de 2008 y enero/junio de 2009. En esas ocasiones se utilizó un modelo especial de examen mixto *–ad hoc–*, con parte de test y parte de preguntas abiertas, necesario para determinar la validez comparando dos procedimientos diferentes de examen. Ese carácter de examen mixto impuso una limitación al número de preguntas de tipo test que podíamos incluir, solamente 30. El estudio de la consistencia interna, que no figuraba en nuestro diseño inicial del trabajo, se inició posteriormente, por lo que ya no permitió utilizar los mismos participantes que en el de validez -salvo en un caso correspondiente a los estudiantes de Enfermería de 2009-; por otra parte, el número de ítems de los test usados para la validez, 30, se consideró un tanto reducido para un adecuado estudio de consistencia interna. Así que el estudio de consistencia interna se hizo en exámenes posteriores y, en su mayoría, a grupos diferentes –salvo en el caso mencionado de Enfermería-. Se escogieron preferentemente los exámenes finales del curso, en mayo o junio, como los adecuados para estudiar la CI, pues el número de preguntas de éstos, en torno a 45, resultaba adecuado; ni eran pocas –como eran las 30 de los exámenes utilizados en el examen

para la validez- ni tampoco un número excesivo que obligase a un gran esfuerzo para introducir los resultados. Hay que tener en cuenta que el estudio de la consistencia interna obliga a introducir en una tabla Excel cada nota obtenida por los alumnos en cada pregunta del test por dos procedimientos diferentes de corrección; se tuvieron que introducir por este procedimiento en total 22358 datos solo para el estudio de CI, lo que aconsejó economizar los esfuerzos buscando exámenes que no tuvieran ni demasiados ítems ni tampoco excesivamente pocos (hubo un grupo, el de Fisioterapia de 2014, en donde se utilizó un examen de 60 ítems, simplemente porque en esa titulación no utilizábamos exámenes con menos preguntas). La encuesta se diseñó en 2012 y se realizó en 2013 y 2014; necesariamente habría de hacerse en el mes de febrero o marzo, pues tenía que ser dos meses posterior al examen de diciembre/enero; eso condicionó los grupos y fechas de hacerla. No hubo coincidencia con los grupos utilizados en el estudio de validez, hecho 4 años antes, ni con los del estudio de consistencia interna, hechos en los exámenes finales por los motivos que antes hemos explicado.

Para el estudio de validez se estudian los exámenes de tipo mixto, TBC y preguntas abiertas, realizados en 5 grupos de alumnos entre los años de 2008 y 2009. Dichos alumnos pertenecían a los cursos 2º y 3º de las titulaciones de Enfermería y Fisioterapia de la Escuela de Enfermería y Fisioterapia perteneciente a la Universidad Pontificia Comillas. Se incluyen todos los sujetos participantes de cada grupo que hicieron el examen. Los cursos y el número de estudiantes participantes en cada uno se exponen en la tabla 3.1.

TABLA 3.1.

Individuos y grupos que participan en el estudio de validez.

Año y curso	N.º de sujetos	mujeres	varones	Tipo de examen
2008 Fisioterapia 3.º	46	31	15	TBC + preguntas abiertas
2009 Fisioterapia 3.º (1)	41	25	16	TBC + preguntas abiertas
2009 Fisioterapia 3.º (2)	44	27	17	TBC + preguntas abiertas
2009 Fisioterapia 2.º	42	25*	13	TBC + preguntas abiertas
2009 Enfermería 3.º	42	34	8	TBC + preguntas abiertas
TOTAL	215	143*	71	

Nota. Los dos exámenes de Fisioterapia 3º de 2009, señalados con los números 1 y 2, se realizan en el mismo grupo de alumnos, pero corresponden a dos exámenes diferentes; el primero hecho en enero y el segundo en junio. *Uno de los estudiantes no pudo ser clasificado ni como mujer ni como varón.

3.2.2. PARTICIPANTES INCLUIDOS EN EL ESTUDIO DE FIABILIDAD

Se estudian los exámenes de tipo TBC realizados en 5 grupos de alumnos entre los años de 2009 y 2014. Dichos alumnos pertenecían a los cursos 2º y 3º de las titulaciones de Enfermería y Fisioterapia de la Escuela de Enfermería y Fisioterapia perteneciente a la Universidad Pontificia Comillas. Se incluyen todos los sujetos participantes de cada grupo que hicieron el examen. Los cursos y el número de estudiantes participantes en cada uno se exponen en la tabla 3.2.

TABLA 3.2.

Participantes en el estudio de fiabilidad y número de preguntas en cada examen.

Año y curso	N.º de sujetos	mujeres	hombres	N.º de ítems por examen
2009 Fisioterapia 2.º	25	14*	10	44
2009 Fisioterapia 3.º	42	26	16	42
2009 Enfermería 3.º	40	33	7	32
2013 Enfermería 2.º	107	97	10	45
2014 Fisioterapia 2.º	37	20	17	60
TOTAL	251	190	60	44.7 de media

Nota. *En uno de los sujetos, del curso de 2º de Fisioterapia de 2009, no se pudo determinar el sexo.

Los exámenes que realizaron los participantes eran los exámenes oficiales de cada materia. El número de preguntas, como es lógico en exámenes de materias y contenidos distintos, era diferente en cada uno de ellos. Algunos exámenes eran combinados, siendo el TBC una parte del examen e incluyendo una parte práctica que no se incluye, naturalmente, en el presente estudio. Todos los exámenes analizados eran de tipo TBC con el mismo formato de preguntas. El procedimiento que se siguió se explica en la sección correspondiente.

Todos los participantes fueron alumnos de la escuela de Enfermería y Fisioterapia de Comillas, de 2º o tercer curso, y pertenecientes a grupos de los que el autor de este estudio era profesor de alguna asignatura. No fue posible disponer de alumnos de otras titulaciones distintas o de otros centros universitarios debido a que no se pudo convencer a otros profesores para introducir el sistema de examen TBC. Además de la resistencia de muchos docentes a modificar su forma de examinar, el TBC es un sistema exigente que requiere información previa y entrenamiento de los alumnos, que tienen que realizar varios exámenes previos para adquirir destreza antes de enfrentarse a una prueba oficial. Para el profesor también supone un esfuerzo mayor que el que exigen otros sistemas de test, pues este procedimiento de examen es más exigente que los test tradicionales en cuanto al buen diseño de los ítems, no admite defectos en su redacción, y no teníamos en nuestro centro un sistema de corrección mecanizada del TBC, obligando al profesor a hacer la corrección manual; una corrección que, además, es algo más compleja que la tradicional, al manejar seis puntuaciones alternativas por cada pregunta. Por todo ello, siendo el único profesor en utilizar este tipo de examen, hubo que hacer el estudio en los alumnos que teníamos a nuestra disposición, los de Enfermería y Fisioterapia de nuestra escuela.

3.2.3. PARTICIPANTES Y GRUPOS EN LAS ENCUESTAS

A los participantes se les pasó un cuestionario de 58 ítems compuesto por una encuesta y dos escalas. La encuesta de 24 ítems, fue de elaboración propia y versaba sobre su opinión respecto a los diversos tipos de examen; las dos escalas fueron: la de enfoques de aprendizaje RSPQ-2F de 20 ítems y una pequeña escala de 5 ítems sobre autorregulación. A esos 49 ítems se les añadieron 9 preguntas más sobre las variables socio-demográficas de los alumnos –edad, género, trabajo-, resultados en exámenes previos y expectativas docentes).

Las encuestas fueron realizadas en 2013 y 2014 en cuatro grupos de enfermería y dos de fisioterapia, respondiendo a ellas un total de 272 alumnos; 133 -de un total de 149 alumnos matriculados-responden a la encuesta el año 2013, y 139 -de un total de 152-responden a la encuesta en 2014.

La edad media de los participantes fue de 21 años y 8 meses (21.7 años, con desviación estándar 3,7 años); respondieron a esta pregunta 261 alumnos de los 272.

El 30.36 % de los estudiantes encuestados declararon ser trabajadores. Definimos ser trabajador como haber realizado algún trabajo retribuido, aunque fuera de forma intermitente o esporádica, desde el comienzo del curso –los 6 meses anteriores a la realización de la encuesta–.

De los 271 participantes que especificaron su género, 55 (el 20.3 %), eran hombres y 216 (79.7%) mujeres. En nuestros alumnos hay un predominio femenino que, por otro lado, es frecuente entre los estudiantes de las Ciencias de la Salud. Hay una marcada diferencia en cuanto al género entre las dos titulaciones; mientras que en

Enfermería solo son del sexo masculino el 11.44 % de los estudiantes –23 de un total de 201 alumnos–, en Fisioterapia el porcentaje de varones llega al 45.7 % –32 varones y 38 mujeres–. La Enfermería es una carrera que tiene tradicionalmente un gran predominio femenino, pero la proporción de los dos géneros es mucho más equilibrada en Fisioterapia. Suele ser habitual en nuestra escuela un porcentaje de estudiantes varones de poco más del 10 % en Enfermería y próximo al 40–50 % en Fisioterapia. La escasez relativa de estudiantes varones supone una limitación a la hora de realizar estudios cuantitativos que analicen las diferencias por géneros; el número de varones es demasiado reducido y dificulta poder observar diferencias estadísticamente significativas si estas no son muy grandes. La distribución de los sujetos incluidos en el estudio según años y carreras la exponemos en la tabla 3.3.

TABLA 3.3.

Número de participantes en cada grupo y año en la encuesta.

Enfermería 2º curso 2013	98 sujetos
Fisioterapia 2º curso 2013	35 sujetos
Enfermería 2º curso 2014	102 sujetos
Fisioterapia 2º curso 2014	37 sujetos
Total, Enfermería y Fisioterapia 2013 – 2014	272 sujetos

Nota. Se incluyen todos los estudiantes de cada grupo que accedieron a realizar la encuesta.

Con la finalidad de reducir el sesgo que el efecto de novedad del examen pudiera provocar en los alumnos, estos tuvieron un entrenamiento previo en el sistema de examen de tipo test basado en el grado de confianza. Ese entrenamiento previo es necesario hacerlo siempre con el TBC, con independencia del presente trabajo; los alumnos no deben enfrentarse al primer examen oficial, con trascendencia para sus calificaciones, sin haber adquirido una cierta práctica del examen basado en el grado de confianza. Nosotros no modificamos, por efecto del estudio, la pauta de entrenamiento que habíamos establecido como una rutina desde que empezamos a utilizar el TBC;

únicamente cuidamos de que fuera exactamente la misma en todos los grupos, con el mismo número de exámenes previos y en fechas parecidas. Durante los 3 meses anteriores al examen, los estudiantes hicieron tres pequeños exámenes, cuya extensión oscilaba entre las 6 y las 15 preguntas, de tipo test con el sistema TBC. Esos exámenes les fueron devueltos a los alumnos tras ser corregidos por el profesor, con indicaciones concretas respecto a su «puntuación en confianza», su tendencia a asumir más o menos riesgos, su acierto al escoger el nivel de confianza en cada pregunta, la nota que obtuvieron y la que hubiesen podido obtener si su elección del nivel de seguridad hubiese sido óptima. Antes de tener que pasar su primera evaluación con el TBC los alumnos deben ser siempre entrenados en el sistema de tal forma que puedan habituarse al mismo y sean conscientes de la incidencia que tiene en su puntuación la utilización de diferentes estrategias de respuesta. Especialmente, es necesario que se percaten de la importancia de reconocer su auténtico nivel de conocimientos, y del hecho, que comprueban al recibir posteriormente el *feed-back* de sus exámenes, de que la mejor estrategia es la sinceridad a la hora de escoger su opción de seguridad. La utilidad del entrenamiento la hemos constatado que al ver cómo en los sucesivos exámenes su nivel de destreza, su habilidad para detectar cuáles son sus conocimientos reales y adoptar la estrategia de confianza adecuada, aumentan considerablemente.

Un ejemplo de la mejora en la habilidad de los estudiantes para escoger su nivel adecuado de seguridad que se produce con la repetición de exámenes pudimos observarlo en el estudio preliminar, y lo exponemos en la tabla 3.4, donde recogemos la modificación en el porcentaje de aciertos con cada nivel de seguridad a medida que se repiten los exámenes y los alumnos van adquiriendo experiencia con ellos. En ella comparamos el porcentaje de respuestas correctas en cada nivel de confianza entre un

primer examen, hecho cuando los alumnos aún no habían tenido contacto previo con el sistema de TBC, y el siguiente examen, tras la experiencia del primero.

TABLA 3.4.

Modificación del porcentaje de respuestas correctas en cada nivel de seguridad en dos exámenes sucesivos.

	totalmente seguro	bastante seguro	No seguro
PRIMER EXAMEN respuestas correctas	79.8 %	75.5 %	52.6 %
SEGUNDO EXAMEN respuestas correctas	96.9 %	87.6 %	50.0 %
Modificación del porcentaje de aciertos entre el 1º y el 2º examen	+ 17.1 %	+ 12.1 %	- 2.6%

Nota. Datos obtenidos de los exámenes de entrenamiento sucesivos realizados en nuestros alumnos en octubre-noviembre de 2008, como parte del estudio de validez.

Como se ve, tras su experiencia previa los alumnos ajustan mucho más su nivel de seguridad a sus conocimientos reales; por eso aumenta su porcentaje de respuestas acertadas con las dos opciones de mayor seguridad y se reduce en la opción de «no seguro».

3.3. INSTRUMENTOS

3.3.1. ESTUDIO PRELIMINAR

Entre 2008 y 2010 el autor desarrolló un estudio sobre el test basado en el grado de seguridad o *test basado en la confianza* (TBC), que fue su trabajo para obtener el Diploma de Estudios Avanzados dentro del ciclo de doctorado de la Universidad Pontificia Comillas. En dicho estudio analizó el funcionamiento de este tipo de examen en alumnos de Enfermería y Fisioterapia. Los objetivos principales eran comprobar qué

tal se adaptaban los alumnos de esas titulaciones en nuestro medio docente a este tipo de exámenes y valorar la validez y fiabilidad de este tipo de evaluación. También se valorarían otros datos complementarios, como eran las diferencias según el género de los alumnos en relación a la elección del nivel de seguridad y la consistencia interna de la puntuación obtenida con este test. Los datos del estudio, que formaron parte del Diploma de Estudios Avanzados del presente doctorado, no han sido publicados y constituyen el estudio preliminar del actual trabajo.

3.3.1.1. FORMATO DE LOS EXÁMENES DE TBC UTILIZADOS POR NOSOTROS

Para nuestro trabajo, tanto en el estudio preliminar de validez y fiabilidad como para la parte principal, centrada en las encuestas, utilizamos los datos obtenidos de los exámenes oficiales de las asignaturas que impartimos. Esos exámenes fueron realizados al finalizar el primer cuatrimestre del curso –diciembre-enero–, o en el final de éste, mayo-junio. En el estudio de validez y en el de consistencia interna utilizamos indistintamente exámenes del primer cuatrimestre o del final del curso. En la encuesta tuvimos que utilizar también datos de los exámenes realizados por el alumno («nota con el TBC» y «nota por aciertos»), aunque esos datos no los obteníamos directamente sino que nos los proporcionaban los participantes; los datos que nos proporcionaban se referían siempre al último examen realizado, que era necesariamente el examen oficial, parcial o final, realizado al finalizar el primer cuatrimestre, en diciembre y enero.

Los exámenes de tipo test basados en la confianza que realizamos nosotros son test de respuesta múltiple, con tres opciones de respuesta. Siempre utilizamos el formato de 3 opciones de respuesta, que es el recomendado también por otros autores (Haladyna

y otros, 2013). Intentamos seguir las recomendaciones que expusimos en la tabla 2.11 del Marco Teórico. No esforzamos en que los enunciados de las preguntas y de las respuestas sean claros, sin ambigüedades, en el mismo tipo de lenguaje en que se imparte la docencia. Procuramos que los distractores sean verosímiles, sin pistas para el alumno; reducimos al máximo las preguntas o respuestas de enunciado negativo y cuando las ponemos, destacamos la negación con diferencias en el tipo de letra. Evitamos siempre incluir la opción «todos correctos», que no funciona adecuadamente en el TBC pues genera dudas innecesarias al alumno –las dudas en este procedimiento penalizan la nota-; por ello es necesario que los distractores sean siempre falsos. Los defectos en la redacción y composición de los ítems, que en otros sistemas de examen de tipo test son perjudiciales porque inducen errores en la respuesta, en el TBC los son especialmente porque, además de lo anterior, reducen la seguridad del alumno, distorsionando la nota.

Un ejemplo real del formato de cada una de las preguntas de nuestro sistema de TBC:

3.- ¿En qué parte del cerebro se regulan la vigilia y el sueño?		
A) Sustancia reticular del tronco encefálico.		
B) Hipotálamo y núcleos de la base del cerebro.		
C) Área cortical límbica del telencéfalo.		
RESPUESTA CORRECTA:		
Totalmente seguro	Bastante seguro	No estoy seguro
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Nos esforzamos en que, ni por la extensión ni por el formato o el lenguaje, haya pistas para el alumno en los distractores, que han de tener apariencia verosímil y poder ser identificados como falsos por el alumno.

Cuando el TBC es el único procedimiento de examen, procuramos que el examen tenga una cifra próxima a 60 preguntas -entre 54 y 60-. Esa es una cifra adecuada de ítems en el formato de test con tres respuestas por pregunta. En general, no somos partidarios de poner exámenes con más de 60 preguntas por la fatiga que inducen en el examinando, que hace que aumente la posibilidad de que cometa errores.

Cuando el test es un componente no mayoritario de la nota de un examen, que incluye otros sistemas de evaluación en el mismo acto, solemos reducir algo el número de preguntas; habitualmente ponemos en torno a 45, cifra que nos parece todavía suficiente para este tipo de examen. En el estudio preliminar, en el que utilizamos exámenes mixtos de test y preguntas abiertas, utilizamos en alguna ocasión solo 30 preguntas de tipo test (que es la cifra mínima para un test de tres opciones de respuesta). Desde el año 2010 siempre utilizamos test con al menos 45 preguntas.

Como dijimos antes, los alumnos ya conocen el sistema de examen, pues se han entrenado en él en tres pequeños exámenes previos. No obstante se les repiten, de forma exhaustiva, todas las instrucciones de nuevo. La tabla de puntuaciones del TBC está expuesta en cada impreso de examen, para que el alumno la tenga presente. Un punto importante que tenemos que recalcar a los estudiantes cuando hacen el examen es que deben responder a todas las preguntas del test, sepan o ignoren la respuesta. Muchos experimentan una gran renuencia a responder por adivinación cuando no saben la respuesta correcta; se les explica que para esa situación está pensada la opción «no seguro», que no descuenta puntuación cuando la respuesta es equivocada. Conseguimos que la gran mayoría de los alumnos responda efectivamente a todas.

En el anexo D exponemos una página de uno de nuestros exámenes (la página pertenece a un examen real de la titulación de Enfermería, que se puso en mayo de 2014). Se ve que nosotros ponemos las casillas con la opción de seguridad tras cada pregunta, en una hoja única que incluye las preguntas y las respuestas. Habitualmente, los demás autores usan una hoja aparte para las respuestas, que puede estar diseñada para corrección mecanizada (Gardner-Medwin 2006). Nosotros usamos un formato que puede resultar más engorroso para la corrección, pero que creemos que disminuye el riesgo de equivocaciones del alumno. Hemos diseñado una hoja de respuestas compatible con la corrección mecanizada, pero no la hemos utilizado en ninguno de los exámenes del presente estudio para minimizar el riesgo de equivocaciones, que siempre se pueden producir cuando la respuesta correcta se traslada desde la hoja de enunciados de los ítems a una hoja aparte de respuesta; lo hemos hecho aceptando que nuestro formato supone una carga de trabajo extra para el profesor.

3.3.1.2. MATERIAS Y ASIGNATURAS EN LAS QUE SE REALIZA EL ESTUDIO DEL TBC

Para el estudio de la validez y consistencia interna del TBC se utilizaron los exámenes oficiales de las asignaturas «Afecciones medicoquirúrgicas», «Enfermería medicoquirúrgica» y «Fisiopatología del paciente crítico». Los datos de las asignaturas y número de participantes del estudio de validez y del de fiabilidad se muestran en las tablas 3.6 –estudio de validez- y 3.7 –estudio de fiabilidad-.

TABLA 3.5.

Individuos y grupos que participan en el estudio de validez.

Año y curso	Asignatura	N.º de sujetos
2008 Fisioterapia 3.º	Afecciones medicoquirúrgicas II	46
2009 Fisioterapia 3.º (enero)	Afecciones medicoquirúrgicas II	41
2009 Fisioterapia 3.º (junio)	Afecciones medicoquirúrgicas II	44
2009 Fisioterapia 2.º	Afecciones medicoquirúrgicas I	42
2009 Enfermería 3.º	Enfermería medicoquirúrgica I	42
TOTAL		215

TABLA 3.6.

Grupo, nombre de la materia y número de sujetos de cada grupo, estudio de fiabilidad.

	ASIGNATURA	Número de sujetos
2009 Fisioterapia 2º	Afecciones medicoquirúrgicas I	25
2009 Fisioterapia 3º	Afecciones medicoquirúrgicas II	42
2009 Enfermería 3º	Enfermería medicoquirúrgica II	40
2013 Enfermería 2º	Fisiopatología de paciente crítico	107
2014 Fisioterapia 2º	Afecciones medicoquirúrgicas II	37
TOTAL estudiantes		251

El contenido de cada asignatura examinada y utilizada en el presente trabajo se expone a continuación.

–Afecciones médico–quirúrgicas I y II en Fisioterapia:

En la materia denominada «Afecciones medicoquirúrgicas», que forma parte del actual grado de Fisioterapia y que formó parte de la antigua diplomatura, el alumno aprende los fundamentos de la patología, tanto médica como quirúrgica, que necesitará para su posterior aprendizaje de las terapéuticas de fisioterapia. Se incluye en el área de Ciencias Básicas dentro de la titulación de Fisioterapia y tiene una ponderación de 12 créditos, 6 correspondientes a Afecciones I y los otros 6 a Afecciones II. El contenido

de la materia es similar en el actual Grado universitario al que tenía cuando estaba incluida en la Diplomatura.

–Enfermería médicoquirúrgica (I y II) en Enfermería:

Es una asignatura que formaba parte de la Diplomatura de Enfermería y que cesó de impartirse en 2011, cuando se transformó en grado universitario. Su contenido consistía en el aprendizaje de los fundamentos de la patología y su aplicación práctica en la actividad enfermera. Esta materia tenía dos partes bien diferenciadas, una de adquisición de conocimientos teóricos sobre patología, que es la que evaluamos en los exámenes de tipo TBC, que es la que hemos utilizado en el presente estudio, y otra de contenidos prácticos, que no se evaluaba por estos procedimientos. La parte de contenidos teóricos estaba ponderada en 9 créditos.

–Fisiopatología del paciente crítico:

Es una materia que forma parte del Grado de Enfermería, donde se imparte en el segundo año. Al igual que la antigua Enfermería Médicoquirúrgica, su contenido es el aprendizaje de la patología humana, centrándose esta asignatura en concreto en las enfermedades que ponen en peligro la vida del paciente, de ahí el nombre de «Fisiopatología del paciente crítico». Forma unidad con otra materia, de carácter práctico, el «Soporte vital», que no se evalúa mediante exámenes. Ambas partes, la

Fisiopatología del paciente crítico y el Soporte vital se reparten seis créditos, 3 para cada una. En el presente estudio incluimos el TBC de la parte de Fisiopatología.

Los exámenes que se analizan son de tipo test con respuesta múltiple; en este caso todas las preguntas tienen 3 opciones de respuesta. Cada examen es corregido de dos maneras, con el sistema de puntuación basado en el grado de seguridad (TBC) y según el número de aciertos y fallos. Utilizamos el sistema de puntuación de Gardner–Medwin (2006) modificado por nosotros. En este sistema, según el grado de seguridad cada alumno puede tener las puntuaciones en cada pregunta que mostramos en la tabla 3.7.

TABLA 3.7.

Tabla de puntuaciones utilizada en nuestros exámenes de TBC

Niveles de seguridad	«totalmente seguro»	«bastante seguro»	«no seguro»
Respuesta correcta	+ 3	+ 2	+ 1
Respuesta incorrecta	-3	-1	0

Nota. Modificación de Valencia sobre el modelo de Gardner–Medwin.

Cada examen es también corregido según el procedimiento habitual, por número de aciertos, en este caso hay solo dos valores posibles (tabla 3.8).

TABLA 3.8.

Tabla de puntuaciones del *test por aciertos*.

	Puntuación
Respuesta correcta	+ 1
Respuesta incorrecta	-0.5

Nota. Penalización según la «fórmula clásica de corrección de la adivinación»

La penalización que se pone por cada pregunta respondida incorrectamente es la que corresponde a un test con 3 opciones de respuesta en cada pregunta según la «fórmula clásica» (ver anexo D). Los alumnos no tienen la opción de dejar la pregunta sin contestar.

La puntuación «oficial» que el alumno recibe en el examen es la que corresponde a la obtenida mediante el TBC; la nota según número de aciertos la utilizamos nosotros exclusivamente para este estudio.

De manera que al alumno se le asignan dos notas en cada examen, una de ellas su nota según el TBC, que puede oscilar de -3 a $+3$ por pregunta, y la otra su nota según número de aciertos; para esta última hay que transformar las 3 notas posibles de la respuesta acertada en la corrección de tipo TBC (3, 2 o 1) en la nota de «1» en el test por aciertos, y las tres notas posibles al fallar la respuesta, según el TBC (-3 , -1 y 0) en «-0.5». Es necesario señalar que el establecimiento de esa doble puntuación, según TBC y según aciertos, no la hemos hecho solo en los exámenes que utilizamos para el estudio de validez y fiabilidad, la hacemos sistemáticamente en todos nuestros exámenes. Posteriormente, informamos al estudiante de ambas puntuaciones. La única puntuación que cuenta para la calificación de la asignatura es la nota de TBC; la nota *por aciertos* es puramente informativa, útil para proporcionar información al estudiante sobre su habilidad con el procedimiento del grado de confianza, comparando la nota obtenida con la que hubiese tenido si se hubiera corregido el examen por el procedimiento tradicional basado solo en aciertos y fallos; así puede mejorar su destreza con el procedimiento de examen en posteriores ocasiones.. También es útil para el profesor conocer ambos tipos de puntuación para poder valorar hasta qué punto el grupo de alumnos domina la técnica del TBC.

3.3.2. INSTRUMENTOS: ENCUESTAS

Como expresamos antes, la encuesta, que consta de 57 ítems, es fundamentalmente la suma de tres encuestas: la encuesta sobre la opinión del alumno frente a los diversos tipos de examen, la encuesta de enfoques de aprendizaje y la encuesta de autorregulación, así como varios ítems más de datos personales del estudiante. La distribución de ítems es la siguiente:

TABLA 3.9.

Características de los datos recogidos en la encuesta

TIPO de DATOS RECOGIDOS	Nº de ítems	Numeración de ítems
Datos personales: edad, género, trabajo.	3	No numerados
Datos académicos: notas(TBC y por aciertos) en el último examen; PC (solo en los alumnos de 2014)	2 (3 en 2014)	No numerados
Resultados y expectativas académicas (escala de 1 a 6)	2	24, 25.
Encuesta sobre los tres tipos de examen (escala de 1 a 6).	25	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28.
Escala R-SPQ-2F, de enfoques de aprendizaje	20	29-48
Escala de autorregulación	5	49-53
TOTAL (datos mixtos, cualitativos, cuantitativos y en escala de 1 a 6)	57 (58 en 2014)	5 (6 en 2014) ítems no numerados. Numerados: 2-53.

Nota: No hay ítem con el número 1.

La encuesta sobre la valoración de diversos tipos de examen –test basado en la confianza, test por aciertos y examen de preguntas abiertas- es de elaboración propia, consta de 25 ítems, agrupados en 8 conjuntos de tres, y recogió la experiencia que adquirimos en una encuesta piloto realizada por nosotros el año anterior. Esa encuesta piloto –realizada en febrero/marzo de 2012- nos permitió seleccionar 21 de los 24 ítems

fue elaborada por nosotros, partiendo de nuestra experiencia con la encuesta piloto que habíamos probado el año anterior. Como se ve en el cuadro anterior, la encuesta recoge 57–o 58 en 2014– ítems divididos en dos partes; en primer lugar datos personales del sujeto, que incluyen la edad, el sexo y la situación laboral; la nota obtenida en el último examen tanto según el grado de seguridad como según el número de aciertos (y en 2014, además la PC); en tercer lugar 52 preguntas o enunciados con las cuales tenía que mostrar el sujeto su grado de acuerdo o desacuerdo en una escala de 1 a 6, siendo 1 «no, nunca, totalmente en desacuerdo» y 6 «siempre, totalmente de acuerdo». En la encuesta se incluyen 24 preguntas sobre las opiniones del alumno sobre tres tipos de examen – test convencional, TBC, preguntas abiertas–y sus hábitos de aprendizaje, los 20 ítems de la encuesta R-SPQ–2F, de enfoques de aprendizaje (Biggs, 2001), con la redacción validada en castellano y en su orden correspondiente, y 5 ítems de autorregulación extraídos del Cuestionario de Aprendizaje Autorregulado (Torre 2006). Se incluyen también tres preguntas aisladas, una sobre la influencia general de los exámenes en el estudio y las otras dos, sobre resultados y expectativas académicas. En el anexo B figura el texto completo de la encuesta, tal como **la recibía** el alumno.

El año anterior, en 2012, se había elaborado una encuesta preliminar, de 41 ítems, como estudio piloto; el objetivo de la encuesta era seleccionar ítems que fueran útiles como instrumento de medida de la influencia de los exámenes sobre el esfuerzo de aprendizaje de los estudiantes. Dicha encuesta se les pasó a los alumnos de 2º curso de Enfermería y Fisioterapia de ese año. Entre los 41 ítems de esa encuesta seleccionamos los que consideramos más adecuados para valorar la influencia del sistema de examen sobre las tareas de aprendizaje del alumno. En total, 22 ítems de los 41 originales se utilizaron en el presente estudio, realizado en 2013 y 2014.

En el diseño de nuestra encuesta, creímos que debíamos limitar el número de preguntas a una cifra que estuviera en el entorno de 50 (sin contar los datos personales y académicos). La razón de esta limitación era el evitarles un esfuerzo excesivo y fatigoso. Al final tuvimos que sobrepasar en algo ese número, llegando a 52 ítems(números 2-53) para abarcar los contenidos que queríamos estudiar. Añadiendo los 5 (o 6 en 2014) datos personales y académicos que se pedía que los participantes cumplimentaras, se llegaba así a la cifra total de 57 ítems en 2013 y 58 en 2014.

3.3.2.1. ÍTEMS DE LA ENCUESTA

Los ítems a los que tenían que responder los alumnos eran:

- Edad (dato numérico)
- Género (dato no numérico –mujer, hombre- que al registrarlo se transformaba en numérico -0, 1-)
- Situación laboral (trabajo/no trabajo, dato no numérico que al registrarlo se transformaba en numérico -1, 0-)
- Puntuación obtenida en el último examen oficial realizado mediante el sistema de «test basado en el grado de seguridad» (dato numérico)
- Puntuación que hubieran obtenido en ese mismo examen si se les hubiera corregido según número de aciertos (dato numérico, a cada alumno se le había proporcionado por escrito previamente la información sobre esas dos notas, TBC y «por aciertos», para poder rellenar correctamente los dos ítems)

- Diferencia entre la nota obtenida según el TBC y según el test por aciertos. Esta nota no figura como ítem de la encuesta, se calcula haciendo la resta de las dos anteriores (TBC-aciertos)
- Puntuación en confianza. Es un dato que se proporcionó a los participantes del año de 2014. No disponíamos de ese dato en 2013.
- La encuesta propiamente dicha, que constaba de 27 ítems, más la escala RSPQ-2F, de 20 ítems, y una escala de autorregulación de 5 ítems (en total 52 ítems entre encuesta y escalas), todas ellas con seis opciones numéricas de respuesta, puntuadas de 1 a 6, siendo 1 «no, nunca, totalmente en desacuerdo» y 6 «Sí, siempre, totalmente de acuerdo» (los ítems podían consistir en preguntas directas como «¿te gusta tal tipo de examen?» o bien frases con las que el alumno se tenía, o no, que identificar; en ambos casos había la misma escala de respuestas, graduada de 1 a 6). El alumno marcaba un pequeño recuadro debajo del número correspondiente, ordenados de menor a mayor de izquierda a derecha (ver anexo E). El orden era siempre el mismo y no había preguntas de respuesta inversa. Se prefirió un número de valores par, sin valor central, para forzar al alumno a definir su posición sin un valor medio «de refugio». La escala no se puso del 0 al 5 sino del 1 al 6 para evitar el carácter aversivo del 0, que podría distorsionar el resultado. Los ítems miden las siguientes variables que se exponen en la tabla 3.10.

TABLA 3.10.

Variables que se recogen en la encuesta

Variables personales del alumno	
Edad y sexo	
Trabajo	
Resultados en el examen	
Número de aciertos en el examen	
Puntuación en el TBC	
PC (obtenida a partir de los dos datos anteriores, solo en 2014)	

Rasgo que se estudia en cada tipo de examen	
Preferencia del alumno por cada examen	Ítems 2, 3, 4
Influencia de cada tipo de examen en el tiempo dedicado al estudio	Ítems 5, 6, 7
Influencia de cada tipo de examen en el esfuerzo de comprensión	Ítems 8, 9, 10
Influencia de cada tipo de examen en el esfuerzo de memorización	Ítems 11, 12, 13
Influencia general de los exámenes en la forma de estudiar	Ítem 14
Validez de cada tipo de examen como prueba objetiva de los conocimientos del alumno	Ítems 15, 16, 17
Validez de cada tipo de examen como técnica de autoevaluación	Ítems 18, 19, 20
Eficacia de cada tipo de examen para potenciar el aprendizaje	Ítems 21, 22, 23
Valoración del estrés que cada tipo de examen provoca	Ítems 26, 27, 28
Autoeficacia del alumno	
Autoeficacia del alumno con el TBC	Ítem 24
Autoeficacia docente	Ítem 25
Escala R-SPQ-2F (test de Biggs)	
Motivo profundo	Ítems 29, 33, 37, 41, 45
Estrategia profunda	Ítems 30, 34, 38, 42, 46
Motivo superficial	Ítems 31, 35, 39, 43, 47
Estrategia superficial	Ítems 32, 36, 40, 44, 48
Escala de autorregulación	
5 ítems de autorregulación	Ítems 49, 50, 51, 52, 53

En esta encuesta, 25 de los ítems recogen la opinión del alumno sobre los exámenes; 24 de ellos comparan tres modelos diferentes de examen: el test convencional con puntuación según el número de aciertos y fallos, al que llamaremos a partir de ahora «test por aciertos»; el test con puntuación basada en el grado de seguridad, el que es objeto del presente estudio, al que desde ahora, por simplificar, nos referiremos por sus siglas «TBC» (acrónimo de «test basado en la confianza»), buscando un término parecido al inglés TBC o «confidence-based test»; y el examen de preguntas abiertas—o de «temas a desarrollar», que es el término utilizado en la encuesta por ser el más familiar al alumno—. Dentro de esta opinión del alumno sobre cada tipo de examen, cada pregunta corresponde a un rasgo distinto que se quiere estudiar, y en cada una de ellas se le pide la valoración de los tres tipos de examen, siempre en el mismo orden: *test por aciertos*, TBC y *examen de preguntas abiertas*. Se quiere recoger la opinión del alumno sobre los siguientes aspectos:

A -Preferencia del alumno por cada tipo de examen (ítems 2, 3 y 4, cada uno correspondiente a un tipo de examen, el 2 corresponde al *test por aciertos o «test convencional»*, el 3 al TBC» y el 4 «preguntas abiertas»)

¿En qué medida te gustan los siguientes sistemas de evaluación?:
ÍTEM 2 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 3 –Test « teniendo que indicar mi grado de seguridad. »
ÍTEM 4 –Exámenes de tipo desarrollar un tema.

B -Valoración del alumno sobre la influencia general que los exámenes tienen en su forma de estudiar (influyen mucho, bastante, poco, etc..., ítem 14).

ÍTEM 14 –Tu forma de estudiar ¿se ve muy influida por el tipo de examen que te van a poner?

C -Valoración del alumno sobre la forma en que cada tipo de examen influye en sus hábitos de estudio, en los siguientes puntos:

- Tiempo dedicado al estudio(ítems 5, 6 y 7)

¿En qué medida cada tipo de examen te obliga a dedicar más tiempo al estudio?
ÍTEM 5 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 6 –Test « teniendo que indicar mi grado de seguridad. »
ÍTEM 7 –Exámenes de tipo desarrollar un tema.

- Esfuerzo para entender los conceptos (ítems 8, 9 y 10)

¿En qué medida cada tipo de examen te obliga a <u>entender bien todo</u> lo que estás estudiando?:
ÍTEM 8 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 9 –Test « teniendo que indicar mi grado de seguridad. »
ÍTEM 10 –Exámenes de tipo desarrollar un tema.

- Esfuerzo para memorizar datos (ítems 11, 12 y 13)

¿En qué medida cada tipo de examen te obliga a aprender más datos de memoria?
ÍTEM 11 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 12 –Test « teniendo que indicar mi grado de seguridad. ».
ÍTEM 13 –Exámenes de tipo desarrollar un tema.

D -Opinión que el alumno tiene sobre la «validez autopercebida» de cada tipo de examen, por su excelencia como sistema de evaluación en:

- Permitir al profesor conocer mejor lo que el alumno sabe (ítems 15, 16 y 17)

¿Qué tipo de examen crees que le permite al profesor conocer mejor lo que el alumno sabe?
ÍTEM 15 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 16 –Test « teniendo que indicar mi grado de seguridad. ».
ÍTEM 17 –Exámenes de tipo desarrollar un tema.

- Permitirle al alumno descubrir mejor sus lagunas formativas (ítems 18, 19 y 20)

¿Qué tipo de examen te hace darte cuenta mejor de tus lagunas formativas?
ÍTEM 18 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 19 –Test « teniendo que indicar mi grado de seguridad. ».
ÍTEM 20 –Exámenes de tipo desarrollar un tema.

- Mejorar el aprendizaje del alumno (ítems 21, 22 y 23)

¿Qué tipo de examen crees que te hace aprender más?
ÍTEM 21 –Test convencional, sin valorar el grado de seguridad.
ÍTEM 22 –Test « teniendo que indicar mi grado de seguridad. ».
ÍTEM 23 –Exámenes de tipo desarrollar un tema.

E -Valoración del estrés autopercebido por el alumno ante cada tipo de examen (ítems 26, 27 y 28).

¿Te sientes muy estresado antes de hacer un examen?
ÍTEM 26 – Sí, si es un test convencional, sin valorar el grado de seguridad.
ÍTEM 27 –Sí, si es un test « teniendo que indicar mi grado de seguridad. ».
ÍTEM 28 – Sí, si es un examen de tipo desarrollar un tema.

F - Valoración de la habilidad del alumno con el TBC y de expectativas de resultados académicos del alumno (ítems 24 y 25).

¿En qué medida responden a tu situación estas frases?
ÍTEM 24 – En los test teniendo que indicar el grado de seguridad, suelo escoger bien el nivel de seguridad de cada pregunta.
ÍTEM 25 –Creo que este curso tendré varios sobresalientes.

Además del estudio de la opinión del alumno sobre los distintos tipos de examen, se le realizan dos encuestas, de enfoques de aprendizaje y de autorregulación:

G - Encuesta de Biggs de enfoques de aprendizaje, R-SPQ-2F, (Biggs y otros, 2001).

Utilizamos la versión de 20 ítems validada para el castellano por J. Carlos Torre (2005).

No se modifica el orden ni la redacción de los ítems. Estos 20 ítems se dividen en dos grupos de 10 ítems, «enfoque profundo», subdividido en «motivo profundo» y «estrategia profunda», y «enfoque superficial», subdividido igualmente en «motivo superficial» y «estrategia superficial». El orden de los ítems de cada grupo es el siguiente:

- *Motivo profundo*, ítems 29, 33, 37, 41 y 45.

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:
ÍTEM 29 – Me doy cuenta de que estudiar me proporciona con frecuencia un sentimiento de profunda satisfacción personal.
ÍTEM 33 – Me parece que cualquier tema puede llegar a ser muy interés ante una vez que te metes a fondo en él.
ÍTEM 37 –Estudiar temas académicos puede ser tan apasionante como leer una buena novela o ver una buena película.
ÍTEM 41 – Trabajo duro en mis estudios porque encuentro los temas interesantes.
ÍTEM 45 – Voy a la mayoría de las clases con preguntas a las que desearía encontrar respuesta.

- *Estrategia profunda*, ítems 30, 34, 38, 42 y 46.

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:
ÍTEM 30 – Al elaborar o estudiar un tema, no me encuentro satisfecho hasta que me he formado mis propias conclusiones sobre él.
ÍTEM 34 – Encuentro interesantes la mayoría de los nuevos temas y empleo tiempo extra intentando obtener mayor información sobre ellos.
ÍTEM 38 –Me hago preguntas a mí mismo sobre los temas importantes hasta que los comprendo completamente.
ÍTEM 42 – Empleo bastante de mi tiempo libre en buscar más información sobre temas interesantes que se han discutido en las diferentes clases.
ÍTEM 46 – Es muy importante para mí echar un vistazo a la mayoría de las lecturas recomendadas que tienen que ver con las clases.

- *Motivo superficial*, ítems 31, 35, 39, 43 y 47.

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:
ÍTEM 31 – Mi objetivo es aprobar el curso haciendo el mínimo trabajo posible.
ÍTEM 35 – Cuando veo que la asignatura no me resulta muy interesante, trabajo lo menos posible.
ÍTEM 39 –Creo que puedo aprobar la mayoría de las evaluaciones memorizando los aspectos clave en lugar de intentar comprenderlos.
ÍTEM 43 – Me parece que no ayuda estudiar los temas en profundidad. Confunde y hace perder el tiempo cuando todo lo que se necesita es un conocimiento por encima de los temas.
ÍTEM 47 – No le encuentro sentido a aprender contenidos que probablemente no caerán en el examen.

- *Estrategia superficial*, ítems 32, 36, 40, 44 y 48.

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:
ÍTEM 32 – Sólo estudio seriamente lo que se da en las clases o lo que está en los programas detallados de las asignaturas.
ÍTEM 36 – Aprendo las cosas repitiéndolas hasta que me las sé de memoria, aunque no las entienda.
ÍTEM 40 –Generalmente limito mi estudio a lo que está específicamente ordenado, porque creo que es innecesario hacer más.
ÍTEM 44 – Creo que los profesores no deberían esperar que los alumnos dedicaran mucho tiempo a estudiar cosas que no van a caer en el examen.
ÍTEM 48 – Me parece que la mejor manera de aprobar es recordar las respuestas de las preguntas que con más probabilidad pueden caer en el examen.

H - Encuesta de 5 ítems sobre autorregulación, escogidos entre los 20 del cuestionario de aprendizaje autorregulado de Torre (Torre 2007). Son los 5 últimos ítems de nuestra encuesta—el 49, 50, 51, 52 y 53—.

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:
ÍTEM 49 – Mientras estudio, me doy cuenta de si voy cumpliendo o no mis objetivos.
ÍTEM 50 – No utilizo siempre los mismos modos para trabajar/aprender los temas, sé cambiar de estrategia.
ÍTEM 51 – Cuando estoy estudiando una asignatura o módulo, sé exactamente qué quiero conseguir o aprender.
ÍTEM 52 – Cuando estoy estudiando, ante las dificultades, hago un esfuerzo mayor o cambio mi modo de estudiar.
ÍTEM 53 – Cuando estoy trabajando una asignatura, intento identificar las ideas y conceptos que no comprendo.

3.3.3. PROCEDIMIENTOS

3.3.3.1. PROCEDIMIENTO PARA EL ANÁLISIS DE LA VALIDEZ DEL TBC EN COMPARACIÓN CON EL SISTEMA DE TEST CONVENCIONAL (POR ACIERTOS)

Los análisis de la validez de los exámenes se realizan relacionando los resultados obtenidos en ellos con otras *variables-criterio*. En nuestro trabajo, para valorar la validez de los exámenes de tipo test, utilizamos como variable-criterio las notas obtenidas en un examen simultáneo, realizado sobre la misma materia, del tipo de *preguntas abiertas*. Este tipo de examen, preguntas abiertas, lo llevábamos haciendo en nuestras asignaturas desde quince años antes, con plena conformidad sobre su calidad como sistema de evaluación, y, al ser un examen simultáneo sobre la misma materia, nos pareció que funcionaría adecuadamente como variable-criterio. La correlación entre preguntas cerradas de respuesta múltiple (test) y preguntas abiertas confirma y pondera la validez de los test, porque ambos tipos de examen –test y preguntas abiertas- tienden a medir básicamente lo mismo («conocimiento»), pero advirtiendo que la correlación con las preguntas abiertas tiene especial importancia, porque son bastantes las

investigaciones en las que se advierte una relación entre buenos resultados en preguntas abiertas (y también en trabajos para hacer en casa) y un enfoque profundo en el estudio (con más reflexión, relacionando con otros temas, etc.); por ejemplo, entre otros y desde distintos enfoques y planteamientos, Scouller, (1998 y 2000), Ritter, (2000), Narloch, Harbin, Calvin y otros (2006), Balch (2007) y Tian (2007).

Comprobamos la validez de nuestro sistema de test, del TBC, comparándola con la de un test convencional basado en el número de aciertos. Hallamos la correlación $-r$ de Pearson- entre la nota obtenida en cada test y la obtenida en el examen de preguntas abiertas. Comparamos ambos índices de correlación y determinamos qué tipo de test tiene mayor correlación con el examen de preguntas abiertas y, en consecuencia, cuál es el más válido. Consideramos que el coeficiente r entre un sistema de test y el examen de preguntas abiertas debería, al menos, ser superior a 0'500 y significativo para que un test fuera válido. También comprobamos si la diferencia en cuanto a validez (correlación con la variable criterio, que es el examen de preguntas abiertas) de los sistemas de test es estadísticamente significativa.

No hacemos dos exámenes de tipo test diferentes (TBC y *test por aciertos*) sino que el mismo examen **lo corregimos de dos maneras diferentes**, según el grado de confianza y según número de aciertos. Dado que comparamos correlaciones, el que en la puntuación por aciertos se use o no la penalización por respuesta equivocada resulta totalmente irrelevante, pues la penalización no modifica la relación lineal de las notas, por lo tanto no modifica el índice r de Pearson. Al utilizar el mismo examen para comparar dos procedimientos de corrección diferentes eliminamos la varianza que habría si comparáramos dos exámenes distintos, uno con cada procedimiento; no se trata de comparar dos muestras relacionadas sino una única muestra.

3.3.3.2. PROCEDIMIENTO PARA EL ANÁLISIS DE LA CONSISTENCIA INTERNA DEL TBC

En el estudio de la consistencia interna del TBC se calcula la consistencia interna o fiabilidad en cada examen y en cada sistema de puntuación –TBC o *nota por aciertos*– utilizando la fórmula de Cronbach:

$$\alpha = N^{\circ} \text{ de ítems} / (N^{\circ} \text{ de ítems} - 1) \times [1 - \sum \text{varianzas ítems} / \text{varianzas totales}]$$

Esta fórmula resulta sencilla de realizar cuando los datos están introducidos en la tabla Excel, usando las funciones estadísticas y matemáticas de dicha tabla. Posteriormente, al haber utilizado el SPSS en el estudio factorial, hemos verificado los datos de consistencia interna que habíamos obtenido de la fórmula anterior con este último programa.

Los resultados se exponen separadamente, comparando el coeficiente α de Cronbach según el TBC y según el nº de aciertos en cada examen.

3.3.3.3. PROCEDIMIENTO SEGUIDO CON LA ENCUESTA Y LAS ESCALAS DE ENFOQUES DE APRENDIZAJE Y AUTORREGULACIÓN

Los resultados de la encuesta se vertieron en una hoja «Excel» para los cálculos correspondientes. En total se recogieron los resultados de 272 individuos que respondieron a la encuesta en su totalidad, correspondientes a los alumnos de los cursos de 2013 y 2014 de 2º de Enfermería y 2º de Fisioterapia. En los ítems 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 26, 27 y 28 disponíamos además de los

datos de la encuesta piloto que realizamos en 2012, con un total de 138 participantes; pero no los incluimos en el presente estudio.

En los datos obtenidos en la encuesta-incluyendo tanto los datos personales de los sujetos, las diversas puntuaciones del examen y los 52 ítems de la encuesta propiamente dicha- calculamos los *estadísticos básicos*, que establecemos en: media, mediana, varianza, desviación estándar, error típico, curtosis, coeficiente de asimetría, rango, máximo y mínimo, utilizando las funciones estadísticas de la Tabla de «Excel» (anexo F). Calculamos también las correlaciones entre todos los ítems, usando el mismo programa «Excel» (anexo G). Con las correlaciones buscamos agrupamientos de ítems homólogos que tuvieran una adecuada consistencia interna, con coeficiente alfa superior a 0.70. Aquellos agrupamientos que encontramos los incluimos también en el cálculo general de correlaciones de la encuesta. Posteriormente hicimos un estudio factorial exploratorio, usando el programa SPSS, con la función de construcción de agrupamientos con rotación ortogonal de componentes, de los 27 primeros ítems de la encuesta, referidos a la opinión del alumno sobre los tipos de examen, sobre los 20 ítems de la escala RSPQ-2F y sobre las 5 preguntas de autorregulación, para confirmar y completar los factores encontrados manualmente (los resultados del análisis de componentes se muestran en el anexo H). Igualmente calculamos la consistencia interna, mediante el coeficiente alfa de Cronbach, de las dos sub-encuestas que incluimos, el test de Biggs en sus cuatro agrupamientos, y el agrupamiento de ítems de *autorregulación* (anexo H).

Con los resultados de los estadísticos básicos y las correlaciones hemos investigado si las hipótesis que motivaban nuestro estudio se cumplían.

3.3.3.4. PROCEDIMIENTOS: CÁLCULO DEL NIVEL DE CONFIANZA DE LOS ESTUDIANTES

Uno de los datos que hemos querido analizar es el nivel de seguridad –o confianza– del alumno al escoger su opción de seguridad en las preguntas del TBC. Lo hemos calculado siguiendo un procedimiento que ensayamos en el trabajo preliminar que hicimos entre 2009 y 2010. Dicho nivel se basaba en obtener una *puntuación de confianza* de los alumnos y un *índice de confianza*, a partir de la relación entre ésta y la nota por aciertos. Posteriormente hemos completado los datos del *índice de confianza*, que es un parámetro que tiene un importante sesgo dependiendo de la puntuación del alumno en el examen, con un parámetro nuevo, [PC – *nota por aciertos*], que no tiene ese sesgo. El IC nos resultó útil para comparar los datos obtenidos en los exámenes de 2013-2014 con los de nuestro estudio preliminar, 2008-2009. La PC – *nota por aciertos* resulta una medida mejor, sin sesgos, de la **confianza media del alumno**, por lo que nos basamos principalmente en ella para establecerla.

3.3.3.5 PUNTUACIÓN EN CONFIANZA (PC)

Llamamos así a una nota o puntuación que expresa la opción media de seguridad escogida por el alumno en las preguntas del examen. Para ello asignamos una puntuación a cada pregunta según el nivel de seguridad escogido, con independencia de que el alumno la hubiera acertado o fallado. La tabla de puntuaciones que usamos, y que en el trabajo previo establecimos como óptima, es la 3.11.

TABLA 3.11.

Puntuación asignada a los tres niveles de confianza

Nivel de confianza	«Totalmente seguro »	«Bastante seguro »	«No seguro »
Puntuación asignada	10	5	0

Nota. La puntuación asignada es independiente de que el alumno haya respondido correctamente o no.

Estas puntuaciones para medir el nivel de seguridad no son arbitrarias, ni meramente intuitivas, están basadas en el análisis del porcentaje de aciertos que los alumnos obtuvieron al escoger cada nivel de seguridad. Las notas asignadas a las tres opciones de confianza, -10 a la seguridad plena, 5 a la seguridad parcial y 0 a la falta de seguridad- resultan bastante intuitivas, pero no las escogimos por intuición, sino basándonos en el porcentaje de aciertos que los alumnos en cada opción de confianza. Para obtener la puntuación apropiada a cada nivel de seguridad, en los exámenes analizados en el trabajo preliminar seguimos el siguiente procedimiento: puntuamos separadamente las preguntas con la misma opción de seguridad; por un lado las preguntas con la opción «Totalmente seguro», por otro las correspondientes a «Bastante seguro» y, finalmente, aquéllas en las que puso «No seguro». En cada pregunta sumábamos 1 punto si la respuesta era correcta y descontábamos 0.5 puntos si la respuesta era incorrecta, penalización que corresponde exactamente a la aplicación de la fórmula clásica de corrección de la adivinación para preguntas con 3 respuestas, como eran las nuestras. Hicimos la media en los 215 exámenes que analizamos en el trabajo preliminar. Obtuvimos la siguiente tabla de puntuación en cada opción de seguridad:

TABLA 3.12.

Puntuación media obtenida por los alumnos con cada nivel de confianza

Nivel de confianza	«Totalmente seguro »	«Bastante seguro »	«No seguro »
Puntuación media en cada nivel de confianza	9.0	6.0	2.8

Nota. 215 exámenes analizados en 2008 y 2009.

Las puntuaciones de la tabla 3.12 se obtienen corrigiendo y asignando nota separadamente a las preguntas en cada uno de los tres niveles de seguridad. Así, si en el examen solo hubieran estado las preguntas del nivel máximo, *totalmente seguro*, la nota media de la clase habría sido de 9.0; y lo mismo respectivamente para los otros niveles (6 para el nivel intermedio y 2.8 para el nivel mínimo de seguridad).

Resulta racional asignar el valor de 0 al nivel mínimo de confianza 8, pues el alumno declara carecer de ella («no estoy seguro»). Si asignáramos la puntuación de 0 al nivel mínimo de confianza, descontando 2.8 puntos de la nota media obtenida en esa opción, debíamos descontar, igualmente, la misma cantidad a la nota de las otras dos opciones –bastante, totalmente...-; entonces este panel de notas se nos transformaba en el siguiente (con la sustracción de 2.8 a cada una), expuesto en la tabla 3.13.

TABLA 3.13.

Puntuación de cada nivel de confianza, corrigiendo el mínimo a «0»

Nivel de confianza	«Totalmente seguro »	«Bastante seguro »	«No seguro »
Puntuación obtenida en el estudio preliminar	6.2 (9.0-2.8 = 6.2)	3.2 (6.0-2.8 = 3.2)	0 (2.8 - 2.8 = 0)

Nota. Para igualar la nota del nivel mínimo a 0 manteniendo la relación lineal entre las tres notas procedemos a restar 2.8 de todas las notas. De esta manera: $10 - 2.8 = 6.2$; $6.0 - 2.8 = 3.2$; $2.8 - 2.8 = 0$

Obtenida esta tabla modificada, en la que se mantienen las diferencias lineales entre las puntuaciones de cada nivel de seguridad, volvimos a aplicar el principio de que las opciones de seguridad tuvieran un valor numérico acorde con su enunciado –motivo por el que habíamos igualado la opción «no seguro» a 0-, pero ahora lo aplicamos al nivel de máxima seguridad. La expresión «totalmente seguro», nivel máximo posible de seguridad, debería corresponder a un 10 en la escala 0 a 10. Al igualar la opción «Totalmente seguro» a 10, que es el valor numérico intuitivo asociado a la palabra

«totalmente», para mantener la relación proporcional de las demás notas -6.2 – 3.2 – 0-, aplicamos la transformación correspondiente, incremento de todas las notas en un 61.3 %, con lo que la escala de puntuaciones en confianza se nos transforma en 10 – 5.2 – 0 (tabla 3.14).

TABLA 3.14.

Puntuación de cada nivel de confianza, igualando el máximo a «10»

Nivel de confianza	«Totalmente seguro »	«Bastante seguro »	«No seguro »
Puntuación obtenida en exámenes del DEA	10 <i>(6.2x10/6.2 = 10)</i>	5.2 <i>(3.2x10/6.2 = 5.16)</i>	0 <i>(0x10/6.2 = 0)</i>

Nota. Para igualar la nota del nivel máximo a 10 manteniendo la proporción entre las tres notas procedemos a multiplicar las notas por el cociente 10/6.2. De esta manera: $10 \times 10 / 6.2 = 10$; $3.2 \times 10 / 6.2 = 5.2$; $0 \times 10 / 6.2 = 0$

Partiendo del hecho de que la nota media de las respuestas de máxima seguridad fue de 9, la nota de la opción de seguridad intermedia fue de 6 y la nota media de la opción de seguridad mínima fue de 2.8, la relación de proporciones entre las tres, si asignamos el valor de 10 a la máxima confianza y de 0 a la mínima, es prácticamente igual a la de la tabla 3.8, la utilizada por nosotros: igualando a 10 la máxima confianza y a 0 la mínima, la opción intermedia tendría una cuantificación de 5.2 (como hemos expuesto en la Tabla 3.13). Expresado de otra manera, **si aplicamos el valor de 10 a la opción máxima de seguridad y de 0 a la mínima, entonces la opción intermedia deberá tener el valor de 5.2, para que las puntuaciones en confianza reflejen exactamente las proporciones reales de conocimiento que por término medio el alumno tiene cuando las escoge.** Redondeando el 5.2 a 5, la escala 10–5–0 representa casi exactamente la relación de proporciones entre las notas reales con cada nivel de seguridad. Esto lo expresamos en la tabla 3.15, que integra las tablas 3.11, 3.12, 3.13 y 3.14.

TABLA 3.15.

Relación entre la nota media (por aciertos) según el nivel de confianza y puntuación en confianza.

Nivel de confianza	Nota media de las respuestas	Nota con el mínimo = 0	Asignando al máximo el 10	Puntuación en confianza
Máximo	9.0	6.2	10	10
Medio	6.0	3.2	5.2	5
Mínimo	2.8	0	0	0

Nota. Integración de las tablas 3.7, 3.8, 3.9 y 3.10. N = 215 alumnos (DEA). La escala expuesta en la cuarta columna (la de la derecha) es la escala utilizada por nosotros para calcular la *puntuación en confianza*. Se aprecia su similitud con la nota media real (corregida) obtenida en cada nivel de seguridad.

Esta *puntuación en confianza* variaba mucho de alumno a alumno, pero la media de todos los alumnos era sorprendentemente similar a la nota media obtenida en los exámenes, tanto de tipo test como de preguntas abiertas. En la tabla 3.16 exponemos la nota media obtenida en los exámenes de 2008 y 2009, utilizados para el cálculo de la validez, según el *número de aciertos* (con penalización estándar por pregunta fallada), en el examen de tipo *preguntas abiertas* y la *puntuación media en confianza*. A pesar de tratarse de tres puntuaciones diferentes, una de un examen de tipo test, la otra de un examen de preguntas abiertas y la tercera una medida de la confianza del alumno al contestar –independiente de su acierto o fallo– las tres se diferencian solo en unas pocas centésimas de punto, solo un 0.5 % de diferencia entre la más elevada (puntuación en autoconfianza) y la más baja (nota por aciertos). Esta correspondencia tan estrecha entre medidas tan diferentes fue inesperada para nosotros y, desde luego, sorprendente.

TABLA 3.16.

Comparación de las notas medias en el test por aciertos y examen de preguntas abiertas y el nivel medio de confianza.

TIPO DE PUNTUACIÓN	Nota media
Nota media en el test según aciertos (puntuación 1, -0.5, por cada pregunta)	7.35
Nota media en preguntas abiertas	7.38
Puntuación media del grupo <i>en autoconfianza</i> (escala 10-5-0)	7.39

Nota. N = 215 alumnos. Exámenes de 2008 y 2009 utilizados para calcular la validez del TBC.

La sorprendente similitud entre la *puntuación en confianza*, que es una valoración del nivel de seguridad medio expresado por el alumno en el examen, y las notas reales obtenidas por dos procedimientos de examen diferentes avalan la consistencia de este índice como valoración de la autoconfianza de los alumnos. Creemos que es una excelente manera de medir este parámetro, y por eso la hemos vuelto a utilizar en el estudio central de esta investigación.

En todas las encuestas hemos pedido a los alumnos que pusieran su nota en el examen, según el TBC y también según el número de aciertos. Los datos que los alumnos nos proporcionaron de sus puntuaciones eran fiables y exactos, pues previamente a la encuesta a cada alumno le habíamos entregado un informe de su examen donde figuraban esas puntuaciones, práctica que seguimos en todos nuestros exámenes, pues consideramos que el alumno debe recibir la máxima información posible sobre su evaluación. De esa manera, el alumno sabía exactamente qué tal le había ido en el examen con el sistema del grado de confianza, si le había beneficiado o perjudicado en relación a la nota que hubiese tenido en un examen convencional y en qué medida exacta lo había hecho. Esa información, que el alumno volcaba después en dos preguntas de la encuesta, nos permitía disponer de datos correctos y exactos sobre sus notas. En la encuesta había una pregunta –el ítem 24- en donde el alumno tenía que responder precisamente a esa cuestión: lo acertado o desacertado que había estado al escoger su grado de confianza en las preguntas; aunque el alumno disponía del dato numérico objetivo sobre su grado de acierto -la diferencia entre su nota con el TBC y la nota por aciertos-, su respuesta a esa pregunta no se basaba tanto en el análisis de esas puntuaciones como en una interpretación subjetiva de su acierto en el examen, como se demuestra más adelante en el estudio de correlaciones (anexo G).

En las encuestas de 2014 pudimos obtener, a partir de los dos resultados del examen del alumno, su «*puntuación en confianza*» y disponer de ese dato para nuestros cálculos. Para ello dispusimos de una tabla en donde los alumnos figuraban con una clave, para preservar el anonimato, y en donde, junto con su número de aciertos o puntuación TBC, figuraba su «*puntuación en confianza*» (a partir de ahora la llamaremos PC). Al no haber habido ningún caso de coincidencia en dos personas de la misma nota TBC y el mismo número de aciertos, pudimos obtener la PC en cada uno de los encuestados de 2014. En total, tuvimos la PC en 139 alumnos (108 mujeres y 32 varones), los de Enfermería y Fisioterapia que rellenaron las encuestas en 2014, menos 2 que no pusieron el dato.

3.3.3.6. ÍNDICE DE CONFIANZA (IC) Y PC–NOTA POR ACIERTOS

En el marco teórico expusimos la estrecha relación entre la *confianza* y la preparación académica del estudiante y, en consecuencia, la nota obtenida en el examen (Morony y otros, 2013), por lo que, para poder comparar a los alumnos entre sí en cuanto a su autoconfianza, había que ponerla en relación con la nota en el test. Utilizar como variable control la nota obtenida por el estudiante en el TBC no sería correcto, pues esa nota depende en gran medida de la autoconfianza del alumno, por lo que es una *variable dependiente* de la PC. Por ello utilizamos la nota *según el número de aciertos*. Así obtuvimos el *índice de confianza*, IC, resultado del cociente entre la PC y la nota por aciertos:

$$IC = PC / (\text{nota por aciertos})$$

Dicho índice tendía a la unidad y se entendía que los valores superiores a la unidad indicaban autoconfianza, tanto más intensa cuanto mayores fueran, y los valores inferiores a «1» indicaban falta de confianza.

Este índice funciona razonablemente bien cuando la nota del alumno es alta o media, pero para exámenes con puntuaciones muy bajas sufre una importante distorsión, al reducirse demasiado el denominador –la nota obtenida por el alumno-. Con notas muy bajas, este índice tiende a hacerse muy elevado. Así, por ejemplo, si uno de los alumnos tuviera de nota por aciertos un 0, su IC tendería a infinito, destruyendo toda posibilidad de cálculo matemático de ese índice.

Por ese motivo, aunque hemos calculado el IC en los 139 participantes que rellenaron la encuesta en 2014, hemos buscado otro parámetro que midiera la confianza media del estudiante y que no sufriera de esa distorsión. La forma más sencilla de calcular la relación de la autoconfianza del alumno con su nota en el examen es la de restarlas. Nosotros restamos de la primera –*puntuación en confianza*, «PC»- la segunda –nota por aciertos, «NA»-, y así obtuvimos un índice que nos indicaba la tendencia a la confianza, o al riesgo, del alumno sin las distorsiones inaceptables del IC. Este procedimiento para calcular la relación entre la confianza y el resultado del examen es similar al utilizado por Stankov y otros (2012), tal como ya hemos expuesto en el capítulo anterior. Pero Stankov resta de la *puntuación en confianza* el número de aciertos, sin penalización por las respuestas erróneas. Su nota para la confianza la obtiene de una encuesta que hace simultáneamente a cada pregunta del examen y en la que el estudiante ha de valorar su seguridad en tener bien la pregunta en una escala de 0 a 10, es decir, una escala de 11 valores. Nuestra escala de confianza es de tres valores - «totalmente seguro», 1 punto; «bastante seguro», 0.5 puntos, y «no seguro», 0 puntos.

Nosotros restamos de la PC la nota por aciertos, que incluye la penalización por respuesta errónea, y nuestro sistema, como se demuestra en la tabla 3.13, resulta estar casi perfectamente ajustado, al ser la puntuación media en confianza de nuestros alumnos (7.35) casi igual a su nota por aciertos (7.39).

A este nuevo índice, semejante al parámetro que mencionamos en el capítulo anterior y que los autores en lengua inglesa denominan «sesgo de confianza» (*bias u overconfidence*) (Stankov y otros, 2012), no le hemos dado un nombre especial, simplemente «*puntuación en confianza – nota por aciertos*», o **PC-NA** en acrónimo. Este será el parámetro en el que nos vamos a basar principalmente para calcular la relación entre confianza y conocimientos del alumno. Como se puede deducir de los datos que expusimos antes, este nuevo parámetro, IC-NA, tiende a cero o a valores muy próximos a cero si la autoconfianza es acorde con el conocimiento real. Un PC-NA positivo indica una elevada autoconfianza con relación a los conocimientos, un PC-NA negativo indica lo contrario.

Por último, nosotros utilizamos un cuarto índice de confianza, el que expresamos como $|PC-NA|$, que es el anterior (PC-NA) en valores absolutos (sin signo negativo). Si el PC-NA lo expresamos en valores absolutos, nos dará una idea de lo que el alumno se ha alejado de la situación ideal, que sería la de que la confianza del alumno y su grado de conocimiento estuvieran lo más cerca posible (y, por tanto, la PC-NA fuera de 0), **con independencia de que sea por exceso o por defecto de autoconfianza**. Este parámetro, $|PC-NA|$, lo podemos considerar una medida de la «sinceridad» del alumno, y de su metacognición –su *automonitorización* de los propios conocimientos- en relación al dominio de la materia; nos permitirá evaluar hasta qué punto el exceso o

defecto de seguridad le perjudican, mediante el análisis de la correlación de la |PC-NA| con la diferencia «nota en TBC» – «nota por aciertos».

3.4. OBJETIVOS E HIPÓTESIS DEL ESTUDIO

3.4.1. OBJETIVOS

Las hipótesis que queremos contrastar derivan de los objetivos de nuestro estudio, que expusimos en 3.1. Resumidos, los objetivos son los siguientes:

- El primer objetivo es el de comprobar si con el sistema de puntuación basado en el grado de confianza mejora la validez y la consistencia interna de este.
- El segundo objetivo es el de valorar la aceptación del TBC por el estudiante, y si el TBC modifica los hábitos de estudio, o *enfoques de estudio*, de alumno, en comparación con los otros sistemas de evaluación (test por aciertos, examen de preguntas abiertas).
- El tercer objetivo es el de analizar la opinión del alumno sobre este examen, en comparación con los otros sistemas, especialmente en lo referido a lo que constituiría el valor fundamental de un examen: *la validez percibida como sistema de evaluación del aprendizaje*.
- El cuarto objetivo es el de estudiar los enfoques de aprendizaje de nuestros alumnos, profundos y superficiales según el RSPQ-2F, y analizar las correlaciones de estos enfoques con los demás ítems de la encuesta.

- El quinto objetivo es ver si el enfoque de aprendizaje del alumno, profundo o superficial, supone una modificación en la puntuación obtenida con el TBC para ver si este examen premia o incentiva un aprendizaje profundo o superficial.
- El sexto objetivo sería el de estudiar la autorregulación del alumno y correlacionarla con los demás ítems, especialmente aquellos que miden el rendimiento académico y la habilidad en el manejo de TBC
- El séptimo objetivo es analizar la autoconfianza del alumno, parámetro que este tipo de examen permite medir, y ver si se correlaciona con los enfoques de aprendizaje, con la autorregulación y con el género de los estudiantes.
- El octavo objetivo sería ver si la adecuación entre el conocimiento autopercebido por alumno, medido según la PC, y su conocimiento real, medido según el número de aciertos en el test, influyen en la nota obtenida con el TBC.
- El noveno objetivo es ver si el género del alumno influye de alguna manera en la nota obtenida con el TBC en comparación con el test convencional, es decir, si el TBC induce un sesgo de género.

Para conseguir esos objetivos diseñamos el presente estudio, tras nuestra experiencia utilizando el TBC como principal sistema de evaluación en nuestras asignaturas durante 5 años y el estudio piloto que realizamos en 2012 para establecer las preguntas de la encuesta. Estos objetivos se concretan en las hipótesis de estudio.

3.4.2. HIPÓTESIS QUE QUEREMOS CONTRASTAR

1 – El test basado en la confianza tiene mayor validez que el test convencional basado en el número de aciertos.

2 – El sistema de examen de tipo test basado en el grado de seguridad (TBC) ofrece resultados con similar o mayor consistencia interna (fiabilidad) que el examen de tipo test tradicional, basado en el número de aciertos.

3 – Los alumnos refieren que sus hábitos de estudio son condicionados por el sistema de evaluación que espera tener.

4 – Los alumnos refieren que el saber que van a ser examinados mediante el TBC le estimula a dedicar mayor tiempo al estudio, frente al examen de tipo test tradicional *por aciertos*.

5 – Los alumnos refieren que el saber que van a ser examinados mediante el TBC les estimula a realizar un mayor esfuerzo de comprensión de los conceptos aprendidos, en comparación con el examen tradicional *por aciertos* o el examen de *preguntas abiertas*.

6 – Los alumnos refieren que el TBC no les genera rechazo, a pesar de su complejidad.

7 – Los alumnos opinan que el TBC es un sistema de examen mejor que el test convencional *por aciertos* o el examen de *preguntas abiertas*. Consideramos que las características que definen a un examen como «mejor» son que proporcione al profesor mayor información sobre los conocimientos del alumno, que al estudiante le resulte más útil para su autoevaluación y que le estimule más en el aprendizaje.

8 – Los alumnos con alta puntuación en enfoques profundos de aprendizaje tienen mejor opinión del TBC que los alumnos con baja puntuación en estos enfoques.

9 – Los alumnos con enfoques profundos de aprendizaje tienden a tener puntuaciones relativas más altas en el TBC que en los *test por aciertos*.

10 – Los alumnos con mejor autorregulación obtienen mejor puntuación en el TBC que en el *test por aciertos*.

11 – Cuando el conocimiento autopercebido por el alumno es igual al conocimiento real- expresado mediante el número de aciertos- el alumno experimenta un incremento de nota con el TBC frente a la obtenida por *número de aciertos*. La falta de concordancia entre el conocimiento real y el autopercebido, y la falta de sinceridad del alumno, se relacionan con menor puntuación en el TBC.

12 – El TBC no tiene sesgo de género. Ninguno de los dos géneros sale beneficiado o perjudicado por este procedimiento de examen.

3.5. PROCEDIMIENTOS: RECOGIDA DE DATOS.

3.5.1. RECOGIDA DE DATOS DEL ESTUDIO DE CONSISTENCIA INTERNA.

Para el estudio preliminar sobre la validez y la fiabilidad de los sistemas de examen se utilizan los exámenes realizados a los alumnos de los siguientes cursos: segundo y tercero de Fisioterapia (diplomatura) durante los cursos 2008–2009, tercero de Enfermería (diplomatura) de 2009, segundo de Enfermería (grado) en el curso 2012–

2013 y segundo de Fisioterapia (grado) en el curso 2013–2014. La tabla con alumnos, cursos y número de preguntas del test la hemos expuesto anteriormente (tablas 3.1 y 3.2).

En los exámenes que se utilizan para el estudio de la validez, que son mixtos, tipo TBC y de preguntas abiertas, se recogen los resultados del alumno en el test corregido según el sistema basado en la confianza y según aciertos y la nota según preguntas abiertas. Las puntuaciones conseguidas por cada alumno con los tres procedimientos se vuelcan en una tabla Excel y se calcula la correlación entre las notas de los tres exámenes mediante la *r* de Pearson.

Los exámenes que se utilizan para el estudio de fiabilidad o consistencia interna se corrigen mediante dos procedimientos, por número de aciertos y según el grado de seguridad, y se registran las puntuaciones obtenidas con cada procedimiento en cada una de las preguntas planteadas al alumno. Las puntuaciones se vierten en una tabla Excel y se calculan, utilizando sus funciones matemáticas, los estadísticos básicos. Entre esos estadísticos básicos, la tabla de Excel nos proporciona las varianzas. A partir de las varianzas de los ítems y las varianzas de los totales, en cada alumno se calcula el índice alfa de Cronbach como medida de consistencia interna o *fiabilidad* (Morales, 2008, pg. 201-207).

3.5.2. RECOGIDA DE DATOS DE LA ENCUESTA

La encuesta y las escalas RSPQ-2F y de autorregulación se pasan a los alumnos de 2.º de Enfermería y Fisioterapia en el mes de febrero o marzo de los años 2013 y

2014, en dos promociones sucesivas de esas titulaciones. Se realizan en los meses de febrero-marzo porque los alumnos tuvieron un periodo de entrenamiento con el TBC durante el otoño anterior y realizaron en diciembre o enero su primer examen oficial con el sistema TBC. A partir del mes de enero los alumnos conocen ya sus resultados en el examen, con los datos que hemos explicado que se les proporcionaban (nota TBC, puntuación por aciertos y –en 2014- puntuación en confianza). Hemos escogido dejar al menos dos meses de intervalo entre la realización del examen y la encuesta, para que las respuestas del alumno no estén demasiado condicionadas por las emociones asociadas al periodo de exámenes y el conocimiento de sus resultados; así, si los alumnos se habían examinado en diciembre, se les pasaba la encuesta en febrero; si se habían examinado en enero, la encuesta procurábamos ponerla en el mes de marzo.

Como juzgamos que debíamos incluir más de 200 estudiantes en la encuesta, no pudimos hacer la toma de datos solamente en un año, 2013 –solo dispusimos de 133 participantes-, sino que tuvimos que repetirla con los alumnos de la promoción siguiente en 2014, alcanzando finalmente la cifra de 272 participantes, que nos pareció suficiente. Ese retraso de un año en la recogida de datos, tuvo, sin embargo, varios efectos positivos. Por un lado permitió incluir en el estudio a nuevos grupos de alumnos aumentando algo la representatividad de los resultados, al no depender de los sesgos de una única promoción de estudiantes. Por otro lado, entre la recogida de datos de 2013 y la de 2014 se publicó el trabajo de Morony y otros, que valoraba la relación entre la confianza y los resultados en matemáticas del examen PISA (Morony y otros, 2013); en dicho trabajo utilizaban una medida de confianza de características similares a la nuestra (la que hemos denominado PC), y obtenían datos que probaban una relación estrecha entre confianza y resultados académicos (como hemos expuesto en el Marco Teórico).

Eso nos animó a incluir el parámetro de confianza en las encuestas de 2014 y ampliar los objetivos de estudio.

Las encuestas fueron anónimas. Para garantizarla exactitud de los datos referentes a las puntuaciones que los alumnos tuvieron en el examen anterior a la encuesta, se les entregó antes de empezar la encuesta una tarjeta que habrían de quedarse, con su nota en el TBC, la *nota por aciertos* y, a los alumnos de 2014 también su puntuación en confianza. Ellos ya conocían las puntuaciones del examen desde un mes antes, pero así garantizábamos que tenían los datos exactos en el momento de la encuesta.

Los grupos y el número de alumnos de cada uno de ellos y del total se expresan en la tabla 3.17.

TABLA 3.17.
Grupos incluidos en la encuesta de nuestro trabajo.

GRUPO	N.º de sujetos	Mujeres	Hombres
2.º enfermería 2013	98	88	10
2.º fisioterapia 2013	35	20	15
TOTAL 2013	133	108	25
2.º enfermería 2014	102	89	13
2.º fisioterapia 2014	37	19	18
TOTAL 2014	139	108	31
TOTAL 2013–2014	272	216	56

En total había 304 alumnos matriculados en los cursos académicos en los que pasamos la encuesta, de los cuales participaron 272 en el estudio. Los 32 restantes no estaban presentes en ninguno de los momentos en que se pasó la encuesta, que se pasó en varios días sucesivos para reducir al máximo la pérdida de participantes por ausencia del aula.

A cada encuesta se le asignó un número clave y los datos se volcaron en hojas de cálculo de Excel. Se dispuso de 7 hojas de cálculo: 4 individuales de los grupos, 2 hojas de cálculo globales de cada año y una hoja de cálculo general de los 272 encuestados.

Se realizaron las siguientes funciones matemáticas:

- Cálculo de estadísticos básicos en cada una de las variables (media, error típico de la media, mediana, varianza, desviación típica, curtosis, coeficiente de asimetría y rango). Se realizan usando las funciones de la tabla Excel. Los resultados de los estadísticos básicos se exponen en el anexo F.
- Cálculo de las correlaciones aplicando el coeficiente «r de Pearson», que figura entre las funciones de la Tabla de Excel.
- Para el cálculo de las diferencias por género utilizamos dos procedimientos. Por un lado asignamos un número a cada sexo-1 a los varones, 0 a las mujeres (estos números fueron elegidos por azar)-, codificando de esta manera el dato cualitativo en un dato numérico; el estudio de correlaciones nos indicaría así si había sesgos de género. Por otro lado, dividimos a los alumnos en dos cohortes, hombres y mujeres, y calculamos los estadísticos de cada una. En este último caso, para ver si las diferencias eran significativas, aplicábamos la *t de Student*.
- Los agrupamientos de ítems según sus correlaciones se hicieron manualmente, siguiendo la sistemática de agrupar ordenadamente los ítems de mayor correlación (Morales, 2008). Posteriormente, para confirmar los datos de los agrupamientos manuales, introdujimos los datos de la encuesta en el programa SPSS, que realiza automáticamente los agrupamientos para un estudio factorial, que hicimos a continuación.

- La consistencia interna de los agrupamientos se calculó también utilizando las funciones de la tabla de Excel. Aunque entre ellas no figura el cálculo del índice *alfa de Cronbach*, si nos proporciona las varianzas de ítems y de los totales, que permiten calcular ese índice en una simple operación matemática (Morales, 2008). Posteriormente, al utilizar el SPSS para hacer el estudio factorial, aprovechamos ese instrumento para verificar que los valores de consistencia calculados mediante la tabla Excel eran correctos.

4. RESULTADOS

4.1. ESTUDIO PRELIMINAR PARA ESTABLECER LA VALIDEZ Y FIABILIDAD DEL TBC

4.1.1. ESTUDIO DE LA VALIDEZ DEL EXAMEN DE TIPO TEST BASADO EN LA CONFIANZA (TBC)

Como explicamos en el capítulo de Metodología, basamos la comprobación de la validez del TBC en la correlación de las puntuaciones obtenidas con este tipo de examen con las obtenidas en el examen de *preguntas abiertas*, que utilizamos como variable criterio. También hacemos la correlación de la puntuación de los mismos exámenes, corregidos por el sistema tradicional basado solo en el número de aciertos. Por último, comparamos ambas correlaciones y vemos qué tipo de examen, el TBC o el *test por aciertos* tiene una mayor validez –entendida como mayor correlación con la nota de las preguntas abiertas- y si la diferencia de correlaciones tiene significación estadística.

4.1.1.1. ADECUACIÓN DEL NIVEL DE SEGURIDAD A LOS CONOCIMIENTOS REALES DEL ALUMNO

Un primer resultado obtenido fue la comparación de la nota obtenida con la corrección basada en la seguridad y la que hubieran tenido en el mismo examen si se hubiera hecho según el sistema convencional, que cuenta el número de aciertos y resta

una puntuación por preguntas incorrectas, según la fórmula clásica de penalización por respuestas erróneas que vimos más arriba. Obtuvimos la siguiente tabla resumida (4.1):

TABLA 4.1.

Puntuación media de nuestros estudiantes con tres sistemas de examen.

Nota media de los alumnos en el TBC	6.45
Ídem en el test según número de aciertos	7.35
Ídem en el examen de preguntas abiertas	7.38

Nota. Datos extraídos de nuestro trabajo para el DEA.

Como se observa en la tabla 4.1, la puntuación media obtenida con el TBC es menor que la obtenida con el test convencional. Eso es lo esperable, porque ese sistema de evaluación supone mayores descuentos por pregunta fallada con seguridad y menores sumas por las acertadas sin estar totalmente seguro. Por ello, nosotros, en la práctica de las asignaturas en las que utilizamos este sistema de evaluación, compensamos esa tendencia a penalizar la nota del TBC aplicando un factor compensador en el baremo de la nota final que compensaba esa disminución de la nota con el TBC; otros autores hacen algo similar (Gardner–Medwin 2007). En nuestro estudio de validez y fiabilidad, sin embargo, usamos las notas sin ninguna corrección, para no introducir artefactos en el resultado. Hay que tener en cuenta que el hecho de que la nota media del test convencional *por aciertos* sea más próxima a la del examen de *preguntas abiertas* que la obtenida mediante el TBC no quiere decir que haya una correlación mayor entre esos dos tipos de examen que entre el TBC y las *preguntas abiertas*; son datos independientes (en realidad la correlación fue mayor, como veremos más abajo, entre el examen utilizado como criterio, el de preguntas abiertas, y el TBC).

Un criterio para valorar lo adecuado de la elección del nivel de seguridad es ver el porcentaje de respuestas correctas o incorrectas en relación con la opción de seguridad

escogida. Con el sistema de puntuación escogido se puede comprobar, como expresamos en las tablas correspondientes al TBC en el capítulo 2 (tabla 2.17), que el sistema beneficiaba al alumno cuando en la máxima opción de seguridad («totalmente seguro») había obtenido más de un 75 % de aciertos, en la opción de nivel intermedio de seguridad de seguridad («bastante seguro») el nivel de aciertos óptimo estaba entre un 50 % y un 75 % y en las de mínima seguridad el nivel adecuado de aciertos era igual o menor de un 50 %.

En el conjunto de los 5 grupos que analizamos, que sumaban 215 sujetos en total, tal como expusimos en el capítulo de Metodología, obtuvimos la tabla 4.2, que nos indica el porcentaje de aciertos en cada uno de los tres niveles de seguridad y la nota que hubieran obtenido puntuando por número de aciertos si solo se contabilizaran las preguntas de ese mismo nivel de seguridad.

TABLA 4.2.

Porcentaje de aciertos con cada nivel de seguridad.

Nivel de seguridad	Respuestas correctas	Nota que correspondería en un examen <i>por aciertos</i>
Aciertos <i>totalmente seguro</i>	93.48 %	9.0
Aciertos <i>bastante seguro</i>	73.14 %	6.0
Aciertos <i>no seguro</i>	52.03 %	2.8

Nota. Muestra de 215 sujetos

Por consiguiente, en los ítems en los que los alumnos escogían la opción de máxima seguridad («*totalmente seguro*»), sus conocimientos en esas preguntas equivalían a obtener un 9.0 de nota, puntuándolas según el número de aciertos y con el descuento típico por respuesta fallada. Cuando los alumnos escogían la opción intermedia de seguridad («*bastante seguro*») su nota correspondiente era de 6.0 en esas preguntas, y cuando los alumnos escogían la opción de menor seguridad («*no seguro*»)

la nota que les correspondería sería de un 2'8. Eso muestra que los alumnos escogen, al responder a las preguntas, su nivel de seguridad con criterios altamente racionales. En realidad, tienden a no asumir excesivos riesgos (como se ve por un porcentaje de aciertos con la opción de mínima seguridad del 52.03 %). Si damos al nivel mínimo de confianza («no seguro») un valor de cero y al nivel máximo («totalmente seguro») un valor de 10, la relación de las puntuaciones que obtuvimos en cada nivel, 9'0, 6'0 y 2'8, se transforma en 10, 5 y 0, respectivamente (5'16 para el nivel medio, para ser precisos). Eso quiere decir que el nivel intermedio de seguridad («bastante seguro») está exactamente a mitad de camino entre «totalmente seguro» y «no seguro» en nuestro sistema de test. En esto nos hemos basado para establecer los valores de la puntuación en confianza, como hemos explicado en el capítulo 3.

En resumen:

Los alumnos escogen su nivel de confianza con criterios racionales, en correspondencia con sus conocimientos reales.

4.1.1.2. VALIDEZ DEL TBC EN COMPARACIÓN CON LA DEL TEST POR ACIERTOS

Los resultados que obtuvimos en el total de nuestros 215 sujetos en este estudio comparativo de correlaciones, correlacionando la nota obtenida con cada tipo de test con la obtenida en el examen de preguntas abiertas, se exponen en la tabla 4.3.

TABLA 4.3.

Correlación de la nota de cada tipo de test con la nota obtenida en el examen de preguntas abiertas

	Correlación con la nota en <i>preguntas abiertas</i>
Test basado en confianza (TBC)	0.705
Test basado en número de aciertos	0.672
Significación de la diferencia	p < 0.05

Nota. La prueba de la t de Student se hizo entre los valores de la r de Pearson obtenidos en los dos test, utilizando el programa SISA, *online statistical análisis*, para 215 sujetos. Dado que utilizamos una sola muestra con dos sistemas de puntuación, la prueba t de Student es adecuada para valorar la significación estadística de la diferencia.

Como se puede ver en la tabla 4.3, la nota obtenida en el TBC tiene una correlación con la nota obtenida en el examen de preguntas abiertas mayor que la nota del test por aciertos, siendo la diferencia pequeña (0.70 frente a 0.67) pero estadísticamente significativa ($p=0.047$, <0.05). Nuestro estudio mostró que el TBC tenía una validez mayor que el test convencional basado en el número de aciertos, con una diferencia ligera pero estadísticamente significativa.

El examen de tipo TBC ha mostrado una validez significativamente superior al test basado en el número de aciertos.

4.1.2 ESTUDIO DE LA CONSISTENCIA INTERNA DE LOS EXÁMENES DE TIPO TEST, COMPARANDO EL TEST CON PUNTUACIÓN BASADA EN EL GRADO DE SEGURIDAD CON EL TEST POR ACIERTOS

Se han analizado cinco exámenes, realizados entre 2009 y 2013, para valorar la consistencia interna del examen de tipo test con puntuación según el grado de seguridad, comparada con el sistema normal de puntuación según número de aciertos.

Cada examen se corregía por los dos procedimientos. La nota de cada ítem (cada pregunta del test) podía oscilar entre -3 y $+3$ en el caso de la corrección según el grado de seguridad -con seis puntuaciones posibles por cada pregunta-, y entre -0.5 y 1 (dos valores) en la puntuación según el número de aciertos (con la penalización estándar por respuesta incorrecta, con tres opciones de respuesta [$-1/(3-1) = -0.5$]).

TABLA 4.4.
Puntuaciones según el nivel de confianza y el test por aciertos

	Puntuación en el TBC	Puntuación en el test por aciertos
Respuesta correcta	3, 2, 1	1
Respuesta incorrecta	0, -1, -3	-0.5

El mayor número de posibilidades de puntuación en cada pregunta, 6 frente a dos, favorece que el TBC obtenga coeficientes alfa de Cronbach de consistencia interna más altos que el test por aciertos -la fiabilidad aumenta al aumentar el número de respuestas de los ítems, en este caso, el número de puntuaciones posibles en cada pregunta (Morales, 2008, pg.213)-, pero ese factor es contrarrestado por el hecho de que el TBC valora simultáneamente dos conceptos-conocimientos y seguridad (aciertos y confianza)- en lugar de solo uno, conocimientos, como hace el test convencional por aciertos; esta última circunstancia, valoración de dos variables distintas -aunque relacionadas en parte-, tendería a disminuir la consistencia interna. Por ello no resulta posible prever, de antemano, cuál de los dos procedimientos será más fiable. Queremos comprobar si el TBC mantiene, al menos, la consistencia interna del test por aciertos. Los resultados obtenidos en el análisis de fiabilidad de los 5 exámenes se muestran a continuación en la tabla 4.5.

TABLA 4.5.

Fiabilidad de los exámenes según el sistema de corrección

Año y curso	Nº de sujetos	N.º de Ítems	α en el TBC	α test por aciertos	t	Signif.
2009 Fisioterapia 2.º	25	44	0.56	0.30	1.18	p>0.05
2009 Fisioterapia 3.º	42	42	0.83	0.83	0.00	p>0.05
2009 Enfermería 3.º	40	32	0.80	0.78	0.31	p>0.05
2013 Enfermería 2.º	107	45	0.82	0.67	3.32	p<0.01*
2014 Fisioterapia 2.º	37	60	0.87	0.85	0.45	p>0.05
TOTAL	251	44.67	-	-	-	-

Nota. La fiabilidad se calcula mediante el índice α de Cronbach. La significación estadística de la diferencia de alfas se hace utilizando el procedimiento de Feldt (1980).

Por lo tanto, el test con puntuación basada en el grado de seguridad tiende a tener mayor fiabilidad que el test basado solo en aciertos en los exámenes analizados, aunque la diferencia solo es estadísticamente significativa en uno de los casos. En cuatro de los cinco casos el coeficiente alfa es más elevado, y solo en uno, el 2º examen, ambos son iguales. De acuerdo con las recomendaciones de varios autores (Thomson, 1994; APA, 2001 –citada por Morales, 2008, pg. 190–), los coeficientes de fiabilidad de muestras distintas se deben presentar independientemente, pues no son sumables ni tiene sentido hacer la media de ellos, pues la fiabilidad no es una propiedad intrínseca del instrumento (el examen, en este caso) sino de los resultados obtenidos al aplicar el instrumento a una muestra determinada (Morales, 2008, p. 190). Por ello, los coeficientes alfa solo se pueden comparar en cada examen, pero no se puede hacer de la suma de los 5. El reducido número de alumnos en la mayoría de las pruebas, de 25 a 42, dificulta que las diferencias puedan alcanzar significación estadística. Debido a ello, el único examen donde las diferencias son estadísticamente significativas (p<0.01) es el examen de enfermería de 2013, en el que participaron 107 estudiantes.

Esto implica que el valorar dos conceptos aparentemente diferentes, como son *conocimientos* y *confianza*, no reduce la consistencia interna del TBC en comparación

con un examen –el test por aciertos- que solo mide conocimiento. Esto establece un fuerte vínculo entre ambos conceptos, que se pondrá de manifiesto en el estudio de correlaciones. Los *conocimientos* y la *confianza* son dos aspectos de un mismo factor, al que podríamos denominar «dominio de la materia». Sin conocimientos no hay, evidentemente, dominio de la materia; sin confianza tampoco.

Por lo tanto, queda demostrado que:

Los exámenes de tipo TBC analizados muestran una consistencia interna igual o superior a la de esos mismos exámenes corregidos solo por el número de aciertos.

4.2. RESULTADOS DE LA ENCUESTA

Tras el estudio preliminar para comprobar la validez y fiabilidad del TBC, expondremos ahora los resultados de la encuesta y sus análisis.

Como dijimos en el capítulo de Metodología, la encuesta que diseñamos es la suma de varias encuestas menores; una sobre la opinión del alumno respecto a los exámenes (25 ítems), una escala sobre enfoques de aprendizaje (RSPQ-2F, 20 ítems), y una pequeña escala sobre autorregulación (5 ítems) y, finalmente, dos ítems más sobre resultados docentes y autoeficacia.

En la encuesta sobre la opinión de los alumnos frente a los diversos tipos de examen, se estudian los siguientes aspectos:

A – Aceptación de este sistema de examen (TBC) por el alumno, comparándolo con la aceptación del examen de test *por aciertos* y del examen de *preguntas abiertas*.

B – Comprobar que los tipos de exámenes esperados por el alumno condicionan la forma de estudio y ver en qué sentido el TBC modifica los hábitos de estudio, siempre comparándolo con otros dos tipos de examen.

C – Ver en qué medida estos tipos de examen -TBC, test *por aciertos* y *preguntas abiertas*- que se comparan son considerados buenos sistemas de evaluación por el alumno. Consideramos que es un buen sistema de evaluación aquél que permite al profesor conocer mejor lo que el alumno sabe, y al alumno autoevaluarse mejor y mejorar su aprendizaje. Compararemos las puntuaciones en estos conceptos de los tres tipos de examen.

D – Ver en qué medida los tres sistemas de examen le generan estrés al alumno.

En la encuesta sobre los enfoques de aprendizaje de los alumnos se estudian los siguientes aspectos:

A – Estudiar los enfoques de aprendizaje, profundo y superficial, de los alumnos.

B – Comprobar si el TBC premia los enfoques profundos frente al test *por aciertos*.

En la encuesta sobre autorregulación de los alumnos se estudian el siguiente aspecto:

A – Ver si los resultados con el TBC se correlacionan mejor con la autorregulación que los del test por número de aciertos.

Otros aspectos que se estudian en la encuesta son:

A – Ver si el género de los alumnos condiciona los resultados con el TBC y si se correlaciona con los ítems de la encuesta.

B – Analizar la autoconfianza de los alumnos, su «tendencia al riesgo », y ver cómo se correlaciona con los demás datos de la encuesta.

C – Analizar si el grado de sinceridad (o reconocimiento de su dominio de la materia) del alumno al escoger su opción de seguridad, tiene reflejo en la nota obtenida con el TBC y con el test *según aciertos*.

Además de estos puntos referidos al TBC, en la encuesta analizaremos los resultados y correlaciones de los enfoques de aprendizaje, de la autorregulación, de la sinceridad y del género con las demás variables del estudio: edad, resultados académicos, expectativas docentes, etc.

4.2.1. RESULTADOS DE LA ENCUESTA PASADA A LOS ALUMNOS EN FEBRERO-MARZO DE 2013 Y 2014

Los participantes en la encuesta, con su distribución por grupos y géneros, los hemos mostrado en la tabla 3.3 del capítulo de Metodología. En total hubo 272 participantes, 200 de ellos estudiantes de Enfermería y 72 de Fisioterapia; 216 mujeres y 56 varones.

Los resultados de todos los ítems de la encuesta se muestran en el anexo F, donde figuran, además, los estadísticos descriptivos de cada uno de ellos. Pasamos a analizar, a continuación, los resultados.

4.2.1.1. ENCUESTA DE LA VALORACIÓN DE LOS ALUMNOS DE LOS DIFERENTES MODELOS DE EXAMEN (ÍTEMS 2, 3 Y 4)

Para estudiar la aceptación del TBC, comparativamente respecto a la que los alumnos tienen de los otros dos tipos de examen –*test por aciertos* y *preguntas abiertas*– incluimos en la encuesta la siguiente pregunta (ítems 2, 3 y 4):

PREGUNTA: «¿En qué medida te gustan los siguientes sistemas de evaluación?: test convencional, test basado en la seguridad, temas abiertos».

Los resultados se reflejan en la tabla 4.6

TABLA 4.6

Pregunta sobre la aceptación de los tres tipos de examen.

	<i>Test por aciertos</i>	TBC	<i>Preguntas abiertas</i>
MEDIA (de 1 a 6)	4.83	3.68	2.56
Error típico	0.06	0.08	0.09
Desviación estándar	1.077	1.405	1.495
Porcentaje sobre la puntuación máxima	77.6 %	53.2 %	29.8 %
Significación de la diferencia con el TBC	p<0.0001	–	p<0.0001

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Se destaca en negrita el valor más alto.

Las diferencias entre las puntuaciones asignadas a los tres tipos de examen por los alumnos eran significativas, como exponemos en la tabla 4.7.

TABLA 4.7

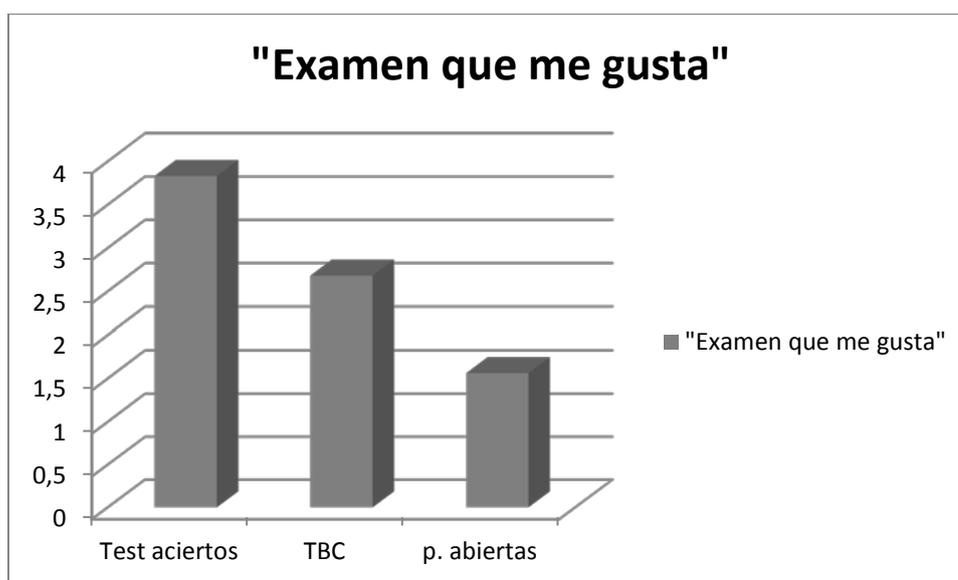
Significación estadística de las diferencias en la aceptación media de cada examen.

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.001
F = 174,39	p<0.001	Eta ² = 0.312

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Para hacer más robusta y exigente la medida de la significación estadística, consideramos que los tres valores dados por el estudiante corresponden a tres medidas sucesivas o repetidas, pues, aunque al valorar los tres tipos de examen responda a los tres ítems a la vez, el alumno tiene que hacer memoria de tres momentos distintos, en asignaturas distintas y con profesores diferentes. En esta situación, y considerando que se cumple la condición de esfericidad, utilizamos la t de Student modificada por los contrastes de Dunn-Bonferroni, que básicamente consisten en dividir el nivel de confianza por el número de comparaciones (3). Posteriormente, ajustamos los valores de p a las nuevas medidas. Esta práctica es considerada por numerosos autores, y por nosotros mismos, exageradamente prudente, pero consideramos que era adecuado realizarla.

La puntuación media asignada por los estudiantes a cada examen, expresada en la tabla 4.3, se visualizan en la figura 4.1.

Figura 4.1



Nota. Para mantener las proporciones reales en las tres barras, el valor de cada examen se ha trasladado linealmente de una escala de 1 a 6 a otra de 0 a 5.

Los alumnos prefieren («les gusta más») el test convencional, *por aciertos*, que los demás sistemas encuestados. A continuación, con un grado intermedia de aceptación, está el TBC y el sistema de evaluación que menos les gusta es el de *preguntas abiertas*.

Aunque menor que el test convencional, el TBC obtiene una buena aceptación; es un sistema de examen que les gusta bastante (puntuación 4), mucho (5) o totalmente (6) al 56.6 %, frente al 43.4 % a los que les gusta poco (3), muy poco (2) o nada (1). Desglosando estos datos obtenemos el siguiente cuadro:

TABLA 4.8.

Aceptación entre los encuestados de TBC.

«Me gusta el TBC»	Número de sujetos y porcentaje
Mucho o totalmente (5 y 6)	86 (31.6 %)
Bastante (4)	68 (25.0 %)
Suma de opiniones positivas (4, 5 y 6)	154 (56.6 %)
Más bien poco (3)	59 (21.7 %)
Muy poco o nada (2 y 1)	59 (21.7 %)
Suma de opiniones negativas (3, 2 y 1)	118 (43.4 %)

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

¿A qué tipo de alumnos les gusta más cada forma de examen? Para responder a esto vamos a estudiar la tabla de correlaciones de esta pregunta (ítems 2,3 y 4). Se utiliza la *r* de Pearson como medida de las correlaciones. Cuando la correlación es estadísticamente significativa se indica debajo de la cifra, entre paréntesis, el nivel de significación. Cuando la correlación no es significativa se señala como «NS», para dar mayor claridad al cuadro. Se destacan en negrilla las correlaciones mayores de 0.202, que corresponden, en una población de 272 sujetos, a un nivel de significación de $p < 0.001$. Los cuadros con todas las correlaciones de los ítems de la encuesta se exponen en el anexo G.

TABLA 4.9

Correlaciones de los ítems de preferencia por exámenes (2, 3, 4) [anexo G]

ÍTEMS Y SU ENUCIADO ABREVIADO	Ítem 2 – «Me gusta el test por aciertos»	Ítem 3 – «Me gusta el TBC»	Ítem 4 – «Me gusta el examen de preguntas abiertas »
EDAD	NS	0.161 (p<0.01)	NS
SEXO	NS	NS	NS
Nota en el TBC	NS	0.349 (p<0.0001)	NS
Número de aciertos	NS	0.305 (p<0.0001)	NS
[TBC– nota por aciertos]	NS	0.168 (p<0.01)	NS
2 – «Me gusta el test por aciertos»	1	-0.402 (p<0.0001)	NS
8 – «El test por aciertos me obliga a entender más cuando estudio »	0.292 (p<0.0001)	-0.153 (p<0.05)	NS
15 – «El test por aciertos le permite al profesor conocer mejor lo que el alumno sabe»	0.218 (p<0.001)	-0.180 (p<0.01)	NS
16 – «El TBC le permite al profesor conocer mejor lo que el alumno sabe »	NS	0.243 (p<0.001)	-0.162 (p<0.01)
17 – «El examen de preguntas abiertas le permite al profesor conocer mejor lo que el alumno sabe »	NS	NS	0.254 (p<0.001)
21 – «El test por aciertos me hace aprender más»	0.244 (p<0.001)	-0.187 (p<0.01)	-0.173 (p<0.01)
22 – «El TBC me hace aprender más»	NS	0.370 (p<0.0001)	NS
23 – «El examen de preguntas abiertas me hace aprender más».	NS	NS	0.348 (p<0.0001)
El test por aciertos es un buen sistema de examen (15+18+21)	0.245 (p<0.001)	-0.175 (p<0.01)	NS
El TBC es un buen sistema de examen» (ítems 16, 19 y 22)	NS	0.309 (p<0.0001)	NS
El examen de preguntas abiertas es un buen sistema de examen (17 + 20 + 23)	NS	NS	0.288 (p<0.001)
24 – «Al hacer el TBC suelo escoger bien mi grado de seguridad »	-0.131 (p<0.05)	0.534 (p<0.0001)	NS
25 – «Este año espero tener varios sobresalientes»	NS	0.199 (p<0.01)	0.152 (p<0.05)

27 – «Me estresa el tener que examinarme mediante el TBC»	NS	-0.165 (p<0.01)	NS
Enfoques PROFUNDOS de aprendizaje	NS	0.225 (p<0.001)	NS
Motivos PROFUNDOS	NS	0.190 (p<0.01)	NS
Estrategias PROFUNDAS	NS	0.215 (p<0.001)	NS
Enfoques SUPERFICIALES de aprendizaje	NS	-0.142 (p<0.05)	NS
Estrategias SUPERFICIALES	NS	-0.168 (p<0.01)	NS
AUTORREGULACIÓN	NS	0.134 (p<0.05)	NS

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes

Los alumnos que tienden a preferir este tipo de examen (el TBC) son:

- Los alumnos que creen haber estado acertados en el último examen de tipo TBC (*correlación alta, r de 0.534*).
- Los alumnos a los que menos les gusta el examen de tipo test *por aciertos* (*correlación moderadamente alta, r de 0.402*)
- Los alumnos que opinan que les hace aprender más y que es un buen sistema de evaluación (*correlación moderada, r de 0.370 y de 0.309, respectivamente*).
- Los alumnos con un mayor nivel de conocimientos, medidos por el examen (*correlación moderada, r de 0.349 si se puntúa según seguridad y r de 0.305 con el número de aciertos*)
- Los alumnos con enfoques profundos de aprendizaje (*correlación moderada, r de 0.225*), más con *estrategias* que con *motivos*.

- Los alumnos que tienen buenas expectativas de resultados académicos (*correlación ligera, r de 0.199*)
- Los alumnos que saben utilizar mejor el TBC (*ligera, r de 0.167*)
- Los alumnos mayores (*ligera, r de 0.161*)
- Los alumnos con mayor autorregulación (*muy ligera, r de 0.134*).
- No hay correlación significativa entre la preferencia por un tipo concreto de examen y el género.

Los alumnos que muestran un mayor rechazo por el TBC frente a los otros tipos de examen, son:

- Aquellos que prefieren el test tradicional por número de aciertos (*moderada-intensa, r de -0.402*).
- Los que creen que el test por aciertos es un buen sistema de examen (*ligera, r de -0.175*).
- Los que tienen enfoques superficiales de aprendizaje, especialmente *estrategias* (*ligera, r de -0.168*).
- Los que manifiestan estresarse con este tipo de examen (*ligera, r de -0.165*).

Los alumnos que manifiestan más preferencia por el *test por aciertos* son:

- Los que manifiestan que el test por aciertos les obliga a entender más lo que estudian (*moderada, r de 0.292*)
- Los que manifiestan que el test por aciertos es un buen sistema de examen (*moderada, r de 0.245*)
- Los que tienen la percepción de dedicar mucho tiempo al estudio con los tres procedimientos de examen (*ligera, r de 0.160*).

Los alumnos que manifiestan preferencia por el examen de *preguntas abiertas* son:

- Los que lo consideran un buen sistema de examen (*moderada, r de 0.288*)
- Los alumnos que tienen buenas expectativas de resultados académicos (*ligera, r de 0.152*).

PERFIL DEL ALUMNO AL QUE GUSTA EL TBC: **Alumnos con mejores resultados académicos que la media, con mayor puntuación en enfoques de aprendizaje profundo, menor en enfoques superficiales, que consideran que saben escoger bien su nivel de seguridad y que se desempeñan ligeramente mejor que la media con el TBC.**

4.2.1.2. INFLUENCIA DE LOS DIVERSOS TIPOS DE EXAMEN SOBRE LOS HÁBITOS DE ESTUDIO DE LOS ALUMNOS

Antes de analizar las modificaciones en los hábitos de aprendizaje, que cada sistema de examen puede inducir en los alumnos, incluimos un ítem (el 14) en el que les preguntábamos si el tipo de examen influía sobre sus hábitos de estudio. El enunciado de este ítem era:

1ª pregunta (Ítem 14): *Tu forma de estudiar ¿se ve muy influida por el tipo de examen que te van a poner?*

Resultados (anexo F):

TABLA 4.10

Influencia del tipo de examen sobre los hábitos de estudio del alumnos

Media (de 1 a 6)	4.60
Desviación estándar	1.539
Porcentaje sobre la puntuación máxima	72.0 %
Nº de sujetos que dieron de puntuación 4, 5 o 6 a este ítem (les influye bastante, mucho o totalmente)	76'8 % (209)
Nº de sujetos que puntuó 1, 2 o 3 a este ítem (les influye nada, muy poco o poco)	23.2 % (63)
Nº de sujetos que opina que el tipo de examen influye nada o muy poco en su estudio (1 y 2)	13.6 % (37)

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes

Los alumnos reconocen que su forma de estudiar se ve bastante influida por la forma de examen que esperan. En total, 209 de los 272 alumnos encuestados en este ítem en 2013 y 2014 (el 76.8 %) respondieron que la forma de estudiar suya se ve bastante, mucho o totalmente influida por el tipo de examen. Así que, de acuerdo con su opinión, mediante el tipo de examen que se les ponga se podría influir en sus hábitos de estudio y aprendizaje.

La respuesta a esta pregunta no se correlaciona con la nota obtenida por el alumno en el examen ni con casi ningún otro ítem de la encuesta. Sólo aparece una correlación significativa, pero baja, con estrategias de estudio superficiales (r de 0.177, $p < 0.01$). Así pues, los alumnos están de acuerdo en que el tipo de examen condiciona su forma de estudiar, y esta idea no se modifica ni con los resultados académicos del alumno, ni con el género o la edad.

En las siguientes preguntas veremos cómo cada tipo de examen influye en la forma de estudiar. Compararemos los resultados obtenidos con el TBC con los obtenidos con los dos otros tipos de examen analizados, el test con puntuación según el número de aciertos y el examen de preguntas abiertas.

2ª Pregunta: ¿En qué medida cada tipo de examen te obliga a dedicar **más tiempo al estudio?**

Ítems 5 (test por aciertos), 6 (TBC) y 7 (preguntas abiertas), resumido de anexo F:

TABLA 4.11

Tiempo dedicado al estudio con cada tipo de examen.

	Test por aciertos	TBC	Preguntas abiertas
MEDIA (de 1 a 6)	4.28	5.02	5.46
Desviación estándar	0.997	0.807	0.897
Error típico	0.06	0.05	0.06
Porcentaje sobre la puntuación máxima	65.8 %	81.0 %	89.2 %
Significación de la diferencia con el TBC	$p < 0.001$	1	$p < 0.001$

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Se destaca en negrita el resultado más elevado de los tres.

Las diferencias entre las puntuaciones medias a los tres exámenes eran estadísticamente significativas, como se ve en la tabla 4.12.

TABLA 4.12

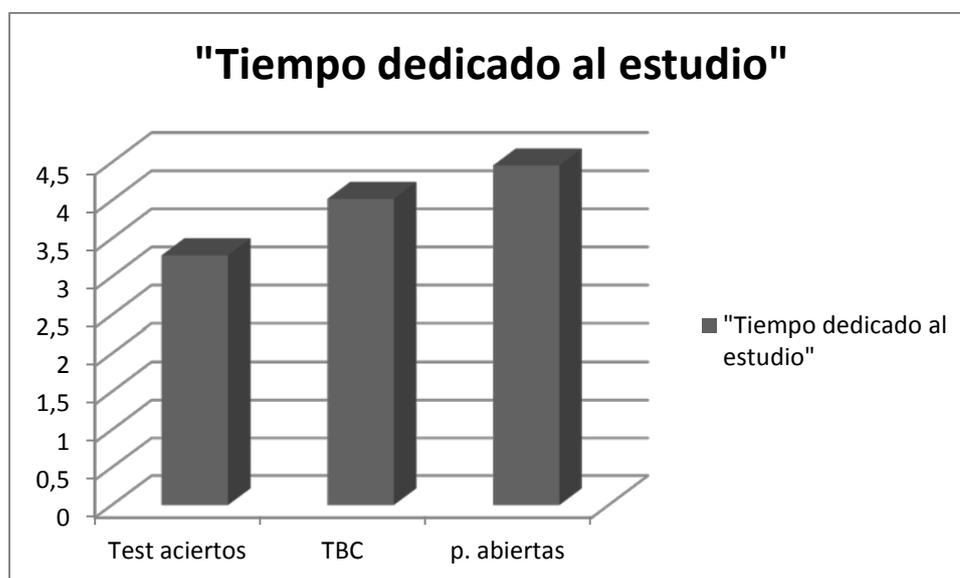
Significación estadística de las diferencias

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.001
F = 166,57	p<0.001	Eta ² = 0.381

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Consideramos que los tres valores dados por el estudiante corresponden a tres medidas sucesivas o repetidas; en consecuencia, utilizamos la t de Student modificada por los contrastes de Dunn-Bonferroni, para 3 comparaciones.

La relación entre las puntuaciones otorgadas por el estudiante a cada examen en tiempo que obliga a dedicar al estudio -expuestas en la tabla 4.8- se aprecian gráficamente en la siguiente figura (4.2)

Figura 4.2



Nota. Para mantener las proporciones reales en las tres barras, el valor de cada examen se ha trasladado linealmente de una escala de 1 a 6 a otra de 0 a 5.

Comparando los datos, se puede observar que el examen de preguntas abiertas es el que, en opinión de los encuestados, les obliga a dedicar más tiempo al estudio, y el de

tipo test basado en el número de aciertos el que menos. En ese sentido, el TBC les obliga a dedicar más tiempo al estudio que el test convencional, aproximándose sus valores a los del examen de tipo preguntas abiertas.

En un análisis detallado de las respuestas sobre el tiempo que dedican a prepararse el examen, el test por aciertos tiene la puntuación más baja: el 76.8 % de los encuestados le asignan la puntuación inferior de los tres tipos, el 50.3 % en solitario, el 26.5 % empatado con otro examen. En cambio, solo le asignan la puntuación más alta en cuanto a trabajo 2 de los encuestados (el 0.6 %), frente a 28 (10.3 %) que asignan esa puntuación al TBC y 136 (50 %) que se la asignan al examen de preguntas abiertas).

TABLA 4.13

Tiempo dedicado al estudio con cada tipo de examen.

Puntuaciones dadas por los alumnos en estos ítems	Test por aciertos	TBC	Preguntas abiertas
6 (máxima puntuación)	21 7.7 %	81 29.8 %	181 66.5 %
5 o 6 (mucho, totalmente)	118 43.4 %	211 77.6 %	231 84.9 %
2 o 1 (poco o nada)	54 19.9 %	11 4.0 %	14 5.1 %

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

Cuando analizamos las correlaciones de estos tres ítems que miden el tiempo que los exámenes obligan al alumno a dedicar al estudio con el resto de la encuesta, encontramos que hay una correlación mayor (siempre significativa) con los siguientes ítems que exponemos en la tabla 4.14.

TABLA 4.14.

Correlaciones de los ítems que valoran el tiempo de estudio (5, 6 y 7) con los demás

	5 el test por aciertos te obliga a dedicar más tiempo al estudio	6 el TBC te obliga a dedicar más tiempo al estudio	7 el EPA te obliga a dedicar más tiempo al estudio
Ítem 6 ¿En qué medida el TBC te obliga a dedicar más tiempo al estudio?	0.488 p < 0.0001	1	0.232 p < 0.001
Ítem 8 ¿En qué medida el test por aciertos te obliga a entender bien todo lo que estás estudiando?	0.370 p < 0.0001	NS	0.149 p < 0.05
Ítem 9 ¿En qué medida el TBC te obliga a entender bien todo lo que estás estudiando?	0.144 p < 0.05	0.282 p < 0.0001	NS
Ítem 10 ¿En qué medida el examen de preguntas abiertas te obliga a entender bien todo lo que estás estudiando?	0.195 p < 0.01	0.164 p < 0.01	0.318 p < 0.0001
Ítem 11 ¿En qué medida el test por aciertos te obliga aprender más datos de memoria?	0.328 p < 0.0001	NS	NS
Ítem 12 ¿En qué medida el TBC te obliga aprender más datos de memoria?	0.245 p < 0.0001	0.400 p < 0.0001	NS
Ítem 13 ¿En qué medida el examen de preguntas abiertas te obliga aprender más datos de memoria?	0.183 p < 0.01	NS	0.240 p < 0.001
Ítem 15 ¿El test por aciertos le permite al profesor conocer mejor lo que el alumno sabe?	0.352 p < 0.0001	NS	NS
Ítem 16 ¿El TBC le permite al profesor conocer mejor lo que el alumno sabe?	0.139 p < 0.05	0.314 p < 0.0001	0.133 p < 0.05
Ítem 17 ¿Las preguntas abiertas le permiten al profesor conocer mejor lo que el alumno sabe?	0.192 p < 0.01	0.154 p < 0.05	0.142 p < 0.05
Ítem 18 ¿El test por aciertos te hace darte cuenta mejor de tus lagunas formativas?	0.233 p < 0.001	NS	0.205 p < 0.001
Ítem 19 ¿El TBC te hace darte cuenta mejor de tus lagunas formativas?	0.176 p < 0.01	0.322 p < 0.0001	NS
Ítem 20 ¿El examen de preguntas abiertas te hace darte cuenta mejor de tus lagunas formativas?	0.128 p < 0.05	0.202 p < 0.001	0.345 p < 0.0001
Ítem 21 ¿El test por aciertos te hace aprender más?	0.368 p < 0.0001	NS	0.145 p < 0.05
Ítem 22 ¿El TBC te hace aprender más?	0.162 p < 0.01	0.283 p < 0.0001	0.202 p < 0.001
Ítem 23 ¿El examen de preguntas abiertas te hace aprender más?	0.235 p < 0.001	0.271 p < 0.001	0.197 p < 0.01

Nota. Ponemos solo los valores de las correlaciones significativas. Destacamos en negrita las correlaciones superiores a 0.20. La tabla completa está en el anexo G.

Solo hemos representado las correlaciones superiores a 0.20 con los tres ítems. Como se ve, hay importantes correlaciones con las demás preguntas que estudian los hábitos de estudio y la opinión sobre los diferentes tipos de examen.

Un hecho muy llamativo en esta pregunta, que interroga sobre el tiempo dedicado al estudio, es que los sujetos que dan una puntuación alta a alguno de los tres ítems, referido a un tipo especial de examen, suelen darla también a los otros dos tipos de examen; es decir, que los alumnos que tienen la percepción de que dedican mucho tiempo al estudio cuando se preparan un examen de un tipo concreto (por aciertos, TBC o preguntas abiertas), también suelen dedicarle mucho tiempo a los exámenes de los demás tipos. Esto se pone de manifiesto en el siguiente cuadro de correlaciones:

TABLA 4.15.
Correlaciones de los ítems de «tiempo de estudio» entre sí.

	Ítem 5 El test por aciertos te obliga a dedicar más tiempo al estudio	Ítem 6 El TBC te obliga a dedicar más tiempo al estudio	Ítem 7 El Examen de preguntas abiertas te obliga a dedicar más tiempo al estudio
6	0.488	1	
7	0.251	0.232	1

Nota. Todas las correlaciones son significativas al menos a un nivel de $p < 0.001$

Estos tres ítems forman, en consecuencia, un cierto agrupamiento. Su consistencia interna (alfa de Cronbach) es de 0.585.

Los alumnos que señalan haber dedicado más tiempo a estudiar cuando esperan tener un TBC son los que piensan que este examen les obliga a comprender bien lo que estudian y a memorizar más datos; igualmente piensan que es un tipo de examen que le permite al profesor conocer mejor lo que el alumno sabe, que a ellos les permite percatarse mejor de sus lagunas formativas y que les hace aprender más. Correlaciones similares se observan con los otros dos tipos de examen.

3ª Pregunta: ¿En qué medida cada tipo de examen te obliga a **entender bien** todo lo que estás estudiando? Ítems 8 (test por aciertos), 9 (TBC) y 10 (preguntas abiertas)

TABLA 4.16.

Medida en la que cada examen obliga al alumno a comprender lo que aprende («esfuerzo de comprensión»).

	Test por aciertos	TBC	Preguntas abiertas
MEDIA (de 1 a 6)	4.68	5.37	4.60
Error típico de la media	0.07	0.05	0.09
Desviación estándar	1.071	0.771	1.350
Porcentaje sobre la puntuación máxima	72.6%	87.0 %	72.0 %

Nota. Los valores completos se exponen en el anexo F. Se destaca en negrita la puntuación mayor.

La significación estadística de las diferencias la exponemos en la tabla 4.17. La diferencia entre la puntuación asignada al TBC y los otros dos tipos de examen es significativa. La diferencia entre la puntuación del test por aciertos y del examen de preguntas abiertas no es, en cambio, significativa.

TABLA 4.17.

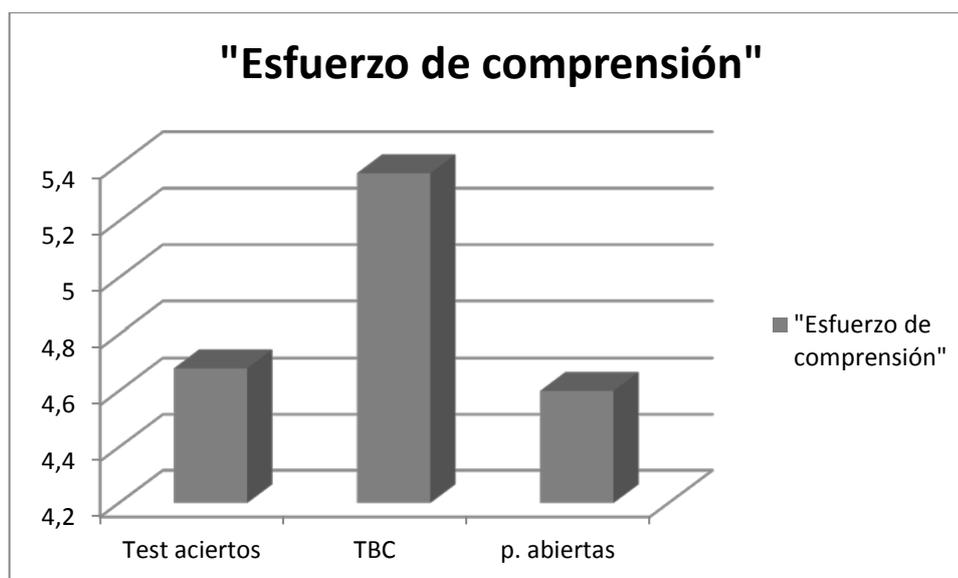
Significación estadística de las diferencias en la media con los tres exámenes.

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	NS	p<0.001
F = 46.29	p < 0.001	Eta ² = 0.147

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Consideramos que los tres valores dados por el estudiante corresponden a tres medidas sucesivas o repetidas; en consecuencia, utilizamos la t de Student modificada por los contrastes de Dunn-Bonferroni, para 3 comparaciones.

Las puntuaciones medias que en esta pregunta los alumnos otorgan a cada examen las podemos apreciar en la figura 4.3.

Figura 4.3



Vemos, por lo tanto, que los alumnos opinan que el TBC les obliga a intentar entender bien lo que están estudiando en mayor medida que los otros dos sistemas de examen, con una diferencia notable y significativa. Cuando se comparan las respuestas a esta pregunta en los otros dos tipos de examen, test *por aciertos* examen de *preguntas abiertas*, se ve que prácticamente no hay diferencias (media de 4.68 y 4.60 respectivamente).

Estudiando las correlaciones de mayor cuantía de esta pregunta obtenemos la información recogida en la tabla 4.18.

¿En qué medida cada tipo de examen te obliga a entender bien todo lo que estás estudiando?

TABLA 4.18.

Tabla de correlaciones de los ítems 8, 9 y 10 con las preguntas que interrogan sobre el constructo que llamamos «*validez autopercebida*» del TBC (ver más adelante).

Preguntas	Ítem 8, <i>test por aciertos</i>	Ítem 9, TBC	Ítem 10, <i>preguntas abiertas</i>
El <i>test por aciertos</i> te hace estudiar más (5)	0.370 (p<0.0001)	0.14 (p<0.05)	0.195 (p<0.01)
El TBC te hace estudiar más (6)	NS	0.282 (p<0.0001)	0.16 (p<0.01)
Las <i>preguntas abiertas</i> te hacen estudiar más (7)	0.15 (p<0.05)	NS	0.318 (p<0.0001)
El <i>test por aciertos</i> hace que el profesor evalúe mejor (15)	0.451 (p<0.0001)	0.13 (p<0.05)	NS
El TBC hace que el profesor evalúe mejor (16)	0.12 (p<0.05)	0.322 (p<0.0001)	NS
Las <i>preguntas abiertas</i> hacen que el profesor evalúe mejor (17)	NS	NS	0.562 (p<0.0001)
El <i>test por aciertos</i> te hace autoevaluarte mejor (18)	0.385 (p<0.0001)	0.16 (p<0.01)	NS
El TBC te hace autoevaluarte mejor (19)	NS	0.270 (p<0.001)	NS
Las <i>preguntas abiertas</i> te hacen autoevaluarte mejor (20)	NS	NS	0.528 (p<0.0001)
El <i>test por aciertos</i> te hace aprender más (21)	0.514 (p<0.0001)	0.198 (p<0.01)	NS
El TBC te hace aprender más (22)	0.18 (p<0.01)	0.369 (p<0.0001)	NS
Las <i>preguntas abiertas</i> te hacen aprender más(23)	NS	NS	0.563 (p<0.0001)

Nota. Destacamos en **negrita** las correlaciones superiores a 0.20. La tabla completa de correlaciones se muestra en el anexo G.

En la tabla 4.18 correlacionamos la puntuación que el alumno da a la pregunta sobre el estímulo a la comprensión de cada tipo de examen con tres preguntas que evalúan lo que viene a ser la opinión del alumno sobre la utilidad de este tipo de examen, a la que denominaremos, a partir de ahora, «*validez autopercebida*» de un sistema de examen: es decir, un sistema de examen que dé una mejor información al profesor de lo que sabe el alumno (ítems 15, 16 y 17), que a éste le haga descubrir mejor sus lagunas formativas (ítems 18, 19 y 20) y que le haga aprender más (ítems 21,

22 y 23) sería el examen que el alumno consideraría más útil; para no usar el término «utilidad», que consideramos inadecuado en un trabajo científico, lo denominamos «validez autopercebida», pero destacando que es un concepto diferente de la *validez* real del examen, que calculamos al principio de este capítulo y que era una cualidad objetiva del sistema de examen que salía del estudio de la correlación de las puntuaciones que se obtenían en él con una variable-criterio; ahora, con el término «*validez autopercebida*» (VA), nos referimos a la opinión del alumno sobre la utilidad de un determinado examen.

De la correlación con estas preguntas que miden la «validez autopercebida» de cada sistema de examen podemos destacar los siguientes hechos: la primera es que los alumnos que opinan que un determinado tipo de examen les obliga a entender mejor los conceptos son los que piensan que ese examen les hace estudiar más (ítems 5.6 y 7), que ese examen permite al profesor evaluar mejor al alumno (15, 15 y 17), que ese examen le permite al alumno autoevaluarse mejor (18, 19 y 20) y que le hace aprender más (21, 22 y 23). Consideramos que esta pregunta, *¿Qué examen te obliga a entender mejor lo que estás estudiando?*, es una pregunta crítica que pone de manifiesto un elemento importante para conceptualizar un examen como bueno: el examen que el alumno considera que le obliga a entender más lo que estudia es el examen mejor (o *más válido*).

<p>El examen que el alumno considera que le obliga a entender mejor lo que estudia es el tipo de examen que él considera que es mejor (o más válido).</p>

En la tabla 4.19 exponemos las correlaciones de los ítems de esta pregunta, hasta qué punto cada tipo de examen obliga a entender bien lo que se estudia, con la escala de

enfoques de aprendizaje y con la de autorregulación, que forman la segunda parte y tercera parte de la encuesta.

TABLA 4.19.

Correlaciones de la pregunta sobre el esfuerzo de comprensión que cada tipo de examen exige con la escala de enfoques de aprendizaje y la de autorregulación.

Pregunta: «esfuerzo de comprensión...»	Test convencional	TBC	Preguntas abiertas
Enfoques profundos (MP + EP)	0.14 (p<0.05)	0.303 (p<0.0001)	0.15 (p<0.05)
Motivo profundo (29+33+37+41+45)	0.13 (p<0.05)	0.316 (p<0.0001)	0.14 (p<0.05)
Estrategia profunda (30+34+38+42+46)	0.11 NS	0.229 (p<0.001)	0.14 (p<0.05)
Enfoques superficiales (MS + ES)	NS	<i>-0.19</i> (p<0.01)	<i>-0.15</i> (p<0.05)
Motivo superficial (31.35.39.43.47)	<i>-0.16</i> (p<0.01)	-0.231 (p<0.001)	<i>-0.13</i> (p<0.05)
Estrategia superficial (32.36.40.44.48)	NS	<i>-0.12</i> (p<0.05)	<i>-0.14</i> (p<0.05)
Autorregulación (49+50+51+52+53)	0.257 (p<0.001)	0.259 (p<0.001)	0.13 (p<0.05)
Pregunta: «esfuerzo de comprensión...»	Test convencional	TBC	Preguntas abiertas

Nota. *r* de Pearson. Destacamos en negrita las correlaciones superiores a 0.20. En *cursiva* las correlaciones inversas. Solo exponemos los ítems que tienen correlaciones significativas.

Esta tabla de correlaciones -4.19- pone de manifiesto que los alumnos que consideran en mayor medida que el TBC es un examen que obliga a comprender lo que se estudia son aquellos que tienen *enfoques profundos de aprendizaje*, sobre todo en el aspecto de *motivos* (r de 0.316) más que en el de *estrategias* (r de 0.229). La correlación de estos enfoques con los otros dos tipos de examen es mucho más débil (todos menores a 0.16). Los enfoques superficiales se correlacionan negativamente, aunque con un coeficiente menor, con la misma opinión sobre el TBC.

Por consiguiente, los alumnos opinan que el examen del tipo TBC les obliga a hacer un esfuerzo mayor para entender mejor lo que estudian que los otros dos tipos de examen. Y esta opinión se mantiene sea cual sea su preferencia por un examen concreto. Los alumnos con enfoques de aprendizaje más profundo son los que consideran en mayor medida que el TBC les obliga a comprender bien todo.

4ª Pregunta: ¿En qué medida cada tipo de examen te obliga aprender más datos de memoria? ítems 11 (test por aciertos), 12 (TBC) y 13 (preguntas abiertas)

TABLA 4.20.

Datos estadísticos de la pregunta sobre el examen que obliga a memorizar más (ítems 11, 12 y 13)

	Test por aciertos (Ítem 11)	TBC (Ítem 12)	Preguntas abiertas (Ítem 13)
MEDIA (de 1 a 6)	4.27	4.79	5.28
Desviación estándar	1.055	1.036	1.039
Error típico de la media	0.07	0.07	0.06
Porcentaje sobre la puntuación máxima	65.4 %	76.4 %	85.4 %

Nota. Datos completos en el anexo F. Se destaca en negrita la puntuación más alta.

Las diferencias entre las medias obtenidas en esta pregunta por los tres tipos de examen son significativas en todos los casos, como se puede ver en la tabla 4.21.

TABLA 4.21.

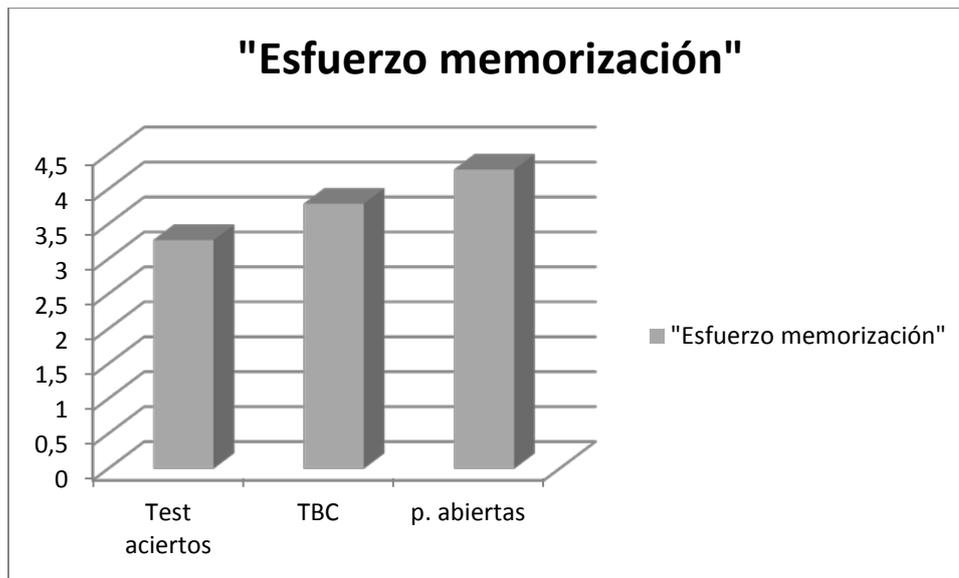
Significación estadística de las diferencias en las medias con los tres exámenes.

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.001
F = 68.35	p < 0.001	Eta ² = 0.201

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes. Consideramos que los tres valores dados por el estudiante corresponden a tres medidas sucesivas o repetidas; en consecuencia, utilizamos la t de Student modificada por los contrastes de Dunn-Bonferroni, para 3 comparaciones.

Las puntuaciones medias que los estudiantes otorgan a cada examen en la cuestión del «esfuerzo de memorización» -expuestas en la tabla 4.20, las mostramos en la figura 4.4.

Figura 4.4



Nota. Para mantener las proporciones reales en las tres barras, el valor de cada examen se ha trasladado linealmente de una escala de 1 a 6 a otra de 0 a 5.

Los alumnos opinan que el examen de *preguntas abiertas* es el que mayor esfuerzo de memorización les obliga a hacer (5.27) y que es *test por aciertos* es el que menos les exige memorizar (3.27). El TBC en este punto está entre los otros dos sistemas de examen (4.82), les obliga a mayor esfuerzo de memorización que el *test por aciertos* pero a menor que el examen de *preguntas abiertas*.

Conclusiones de los ítems que miden la influencia de cada examen en el estudio del alumno:

El test basado en la seguridad (TBC) es el que obliga al alumno a hacer un mayor esfuerzo de comprensión de lo que estudia, por encima del test *por aciertos* y del examen de preguntas abiertas. Obliga al alumno a dedicar más tiempo al estudio y a memorizar más que *el test por aciertos*.

El examen de tipo «preguntas abiertas» es el que les obliga a dedicar más tiempo al estudio y memorizar más datos.

El *test por aciertos* es el que exige del alumno menos tiempo de estudio y menor esfuerzo de memorización. En cuanto al esfuerzo de comprensión, está igualado al examen de preguntas abiertas, ambos por debajo del TBC.

4.2.1.3. ÍTEMS QUE VALORAN LA «VALIDEZ PERCIBIDA POR EL ALUMNO» DE CADA TIPO DE EXAMEN. DESCRIPCIÓN.

Consideramos que el sistema de examen mejor, más útil o *más válido* sería aquél que cumpliera tres características:

- Que proporcionara al profesor un mejor conocimiento del aprendizaje del alumno.
- Que permitiera al alumno autoevaluarse mejor.
- Que estimulara al alumno a aprender más.

Estas tres características las hemos expresado en la encuesta en tres preguntas:

¿Qué tipo de examen crees que le permite al profesor conocer mejor lo que el alumno sabe?

Ítem 15: Test tradicional, sin indicar grado de seguridad.

Ítem 16: Test teniendo que indicar mi grado de seguridad (TBC).

Ítem 17: Examen tipo desarrollar un tema abierto.

¿Qué tipo de examen te hace darte cuenta mejor de tus lagunas formativas?

Ítem 18: Test tradicional, sin indicar grado de seguridad.

Ítem 19: Test teniendo que indicar mi grado de seguridad (TBC).

Ítem 20: Examen tipo desarrollar un tema abierto.

¿Qué tipo de examen crees que te hace aprender más?

Ítem 21: Test tradicional, sin indicar grado de seguridad.

Ítem 22: Test teniendo que indicar mi grado de seguridad (TBC).

Ítem 23: Examen tipo desarrollar un tema abierto.

Estas tres preguntas definen lo que podemos considerar como la *calidad* o, incluso, *excelencia* de un tipo de examen, tal como es percibida por el alumno. Como hemos expresado antes, para este agrupamiento hemos escogido el nombre de «validez percibida por el alumno», o más brevemente «*validez autopercebida*», pues consideramos que expresa la opinión del alumno sobre lo útil, o «válido» que un examen puede ser.

En la primera de las preguntas (ítems 15, 16 y 17) el alumno responde sobre el ajuste de la información que un examen da al profesor con respecto a los conocimientos reales del alumno, algo que él no puede conocer por experiencia directa, sino que ha de deducir comparando su rendimiento en el examen con la autopercepción de sus conocimientos; lo que el alumno está valorando en esta primera pregunta es la validez de cada tipo de examen.

En la segunda y tercera preguntas el alumno sí responde guiándose por su experiencia en la utilización de los distintos exámenes. En el caso del TBC dispone de la experiencia de los varios exámenes de entrenamiento que se han realizado durante los tres meses anteriores.

Vamos a ver los resultados que hemos obtenido con cada una de estas preguntas:

1ª pregunta: ¿Qué tipo de examen crees que le permite **al profesor conocer mejor** lo que el alumno sabe? (ítems 15, 16 y 17).

Los estadísticos básicos de las respuestas de los estudiantes a esta pregunta los exponemos en la tabla 4.22.

TABLA 4.22.

Capacidad de cada examen para informar al profesor de los conocimientos del alumno

Examen que informa mejor al profesor...	Test por aciertos (Ítem 15)	TBC (Ítem 16)	Preguntas abiertas (Ítem 17)
MEDIA (de 1 a 6)	3.97	5.04	4.69
Desviación estándar	1.127	0.979	1.293
Error típico de la media	0.07	0.06	0.09
Porcentaje sobre la puntuación máxima	58.6 %	80.6 %	73.4 %

Nota. Se destaca en negrita el valor más elevado. Los estadísticos completos se muestran en el anexo F.

La significación estadística de las diferencias en la media de los tres exámenes a esta pregunta la exponemos en la tabla 4.23.

TABLA 4.23.

Significación estadística de las diferencias en las medias de los ítems 15, 16 y 17.

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.001
F = 54.99	p< 0.001	Eta ² = 0.170

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

Según la opinión de los alumnos encuestados, el examen que mejor información da al profesor de los conocimientos del examinando es el TBC, seguido del examen de *preguntas abiertas* y, en último lugar, del *test por aciertos*.

2ª pregunta: *¿Qué tipo de examen te hace darte cuenta mejor de tus lagunas formativas?* (ítems 18, 19 y 20)

Los estadísticos básicos de las respuestas de los estudiantes a esta pregunta los exponemos en la tabla 4.24.

TABLA 4.24.

Respuesta a la pregunta sobre la influencia de cada tipo de examen en la autoevaluación del estudiante («darte cuenta de tus lagunas formativas»)

	Test por aciertos (Ítem 18)	TBC (Ítem 19)	Preguntas abiertas (Ítem 20)
MEDIA (de 1 a 6)	4.06	4.95	4.44
Desviación estándar	1.154	1.041	1.355
Error típico de la media	0.07	0.06	0.08
Porcentaje sobre la puntuación máxima	61.4 %	79.6 %	69.0 %
Significación de la diferencia con el TBC	p <0.001		p <0.001

Nota. Se destaca en grinta el valor más elevado. Los estadísticos completos se muestran en el anexo F.

La significación estadística de las diferencias en la media de los tres exámenes a esta pregunta la exponemos en la tabla 4.25.

TABLA 4.25.

Significación estadística de las diferencias en la media de los tres tipos de examen a la pregunta anterior.

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.001
F = 42.27	p < 0.001	Eta ² = 0.135

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

Según la opinión de los alumnos encuestados, el examen que mejor detecta las lagunas formativas del examinando es el TBC, seguido del examen de *preguntas abiertas* y, en último lugar, del *test por aciertos*.

3ª pregunta: *¿Qué tipo de examen crees que te hace aprender más?*

Los estadísticos básicos de las respuestas de los estudiantes a esta pregunta los exponemos en la tabla 4.26.

TABLA 4.26.

Respuesta a la pregunta sobre la influencia de cada examen en «aprender más»

	Test por aciertos (Ítem 21)	TBC (Ítem 22)	Preguntas abiertas (Ítem 23)
MEDIA (de 1 a 6)	4.39	5.07	4.48
Desviación estándar	1.131	0.877	1.396
Error típico de la media	0.07	0.05	0.09
Porcentaje sobre la puntuación máxima	67.8 %	81.4 %	69.6 %

Nota. Se destaca en negrita el valor más elevado. Los estadísticos completos se muestran en el anexo F.

La significación estadística de las diferencias en la media de los tres exámenes a esta pregunta la exponemos en la tabla 4.27.

TABLA 4.27.

Significación estadística de las diferencias en la media de la tabla 4.23

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	NS	p<0.001
F = 29.72	p < 0.001	Eta ² = 0.099

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

Vemos que el TBC es el tipo de examen que recibe la mayor puntuación en los tres ítems, mientras que el *test por aciertos* es el menos puntuado, en la cuestión de cuál es el tipo de examen que estimula más el aprendizaje.

Los alumnos encuestados opinan que el TBC le permite al profesor conocer mejor lo que el alumno sabe, que al alumno le permite detectar mejor sus fallos formativos y que es el que más le hace aprender. Según los criterios que hemos establecido para valorar la bondad de un examen, el TBC es el mejor sistema de evaluación de los tres sometidos a encuesta.

4.2.1.4. ÍTEMS QUE VALORAN LA «VALIDEZ PERCIBIDA POR EL ALUMNO» DE CADA TIPO DE EXAMEN. ANÁLISIS DEL AGRUPAMIENTO Y CORRELACIONES

Las tres preguntas que hemos analizado en el apartado anterior miden una dimensión de los exámenes que consideramos que representa la *validez percibida por el alumno* de un sistema de evaluación, desde tres puntos de vista: la calidad de la información proporcionada al profesor, la proporcionada al alumno y el estímulo para el

aprendizaje. Los resultados obtenidos en cada pregunta de estas tres muestran una alta correlación para cada tipo de examen. Esto nos permite hacer tres agrupamientos:

- Primer agrupamiento: ítems 15, 18 y 21 (*test por aciertos*).
- Segundo agrupamiento: ítems 16, 19 y 22 (TBC).
- Tercer agrupamiento: ítems 17, 20 y 23 (*Examen de preguntas abiertas*).

A estos agrupamientos de tres ítems lo llamaremos: «*opinión sobre la validez del TBC*», «*opinión sobre la validez del test por aciertos*», «*opinión sobre la validez del examen de preguntas abiertas ayuda*», respectivamente al primero, segundo y tercero.

Primer agrupamiento: en el *test por aciertos*, los ítems que valoran su validez para el profesor (15), su utilidad para la autoevaluación (18) y su utilidad para estimular el aprendizaje del alumno (21) muestran el siguiente cuadro de correlaciones:

TABLA 4.28.

Correlación entre los ítems que miden la *validez autopercebida* del *test por aciertos*

	Ítem 15	Ítem 18	Ítem 21
Ítem 15	1		
Ítem 18	0.463	1	
Ítem 21	0.558	0.509	1

Nota. *r* de Pearson. Tabla de correlaciones completa en el anexo G. Todas las correlaciones son significativas ($p < 0.0001$).

El enunciado de estos ítems es el siguiente:

Ítem 15: El test por aciertos le permite al profesor conocer mejor lo que el alumno sabe.

Ítem 18: El test por aciertos te hace darte cuenta mejor de tus lagunas formativas.

Ítem 21: El test por aciertos crees que te hace aprender más.

Este primer agrupamiento, referente a la opinión del alumno sobre el *test por aciertos*, tiene una consistencia interna bastante aceptable, con un coeficiente alfa de Cronbach de **0.756**.

Segundo agrupamiento: estaría formado por los ítems en los que se valora la validez de TBC como sistema de examen. Constaría de los ítems 16 (*el TBC le hace al profesor tener mejor información de mi aprendizaje*), 19 (*el TBC me hace darme cuenta de mis lagunas formativas*) y el 22 (*El TBC me está haciendo aprender más*).

TABLA 4.29.

Correlaciones entre los ítems que miden la *validez autopercebida* del TBC

	Ítem 16	Ítem 19	Ítem 22
Ítem 16	1		
Ítem 19	0.346	1	
Ítem 22	0.605	0.441	1

Nota. *r* de Pearson. Tabla completa en el anexo G. Todas fueron significativas ($p < 0.0001$)

El enunciado de estos ítems es el siguiente

Ítem 16: El TBC le permite al profesor conocer mejor lo que el alumno sabe.

Ítem 19: El TBC te hace darte cuenta mejor de tus lagunas formativas.

Ítem 22: El TBC crees que te hace aprender más.

Este segundo agrupamiento, referente a la opinión del alumno sobre TBC, tiene una consistencia interna también aceptable, con un coeficiente alfa de **0.715**.

Tercer agrupamiento: estaría formado por los ítems en los que se valora la validez de las *preguntas abiertas* como sistema de examen. Constaría de los ítems 17, 20 y 23. Se muestra en la tabla 4.30.

TABLA 4.30.

Correlación entre los ítems que miden la *validez* de las *preguntas abiertas*.

	Ítem 17	Ítem 20	Ítem 23
Ítem 17	1		
Ítem 20	0.480	1	
Ítem 23	0.685	0.591	1

Nota. r de Pearson. Cuadro completo de correlaciones en el anexo G. Todas significativas ($p < 0.0001$)

El enunciado de estos ítems es el siguiente:

Ítem 17: El examen de preguntas abiertas le permite al profesor conocer mejor lo que el alumno sabe.

Ítem 20: El examen de preguntas abiertas te hace darte cuenta mejor de tus lagunas formativas.

Ítem 23: El examen de preguntas abiertas por aciertos crees que te hace aprender más.

Este tercer agrupamiento, referente a la opinión del alumno sobre TBC, tiene una consistencia interna bastante grande, la mayor de este grupo, con un coeficiente alfa de Cronbach de **0.809**.

En los tres casos, la correlación de la pregunta que valora la opinión general del alumno sobre si un tipo de examen le hace aprender más con las otras dos preguntas (*El examen X le permite al profesor conocer mejor lo que el alumno sabe y el examen X le*

hace al alumno darse cuenta mejor de sus lagunas formativas) da un índice *r* de Pearson alto en cada tipo de examen, confirmado la estrecha relación que hay entre la apreciación de que el sistema de examen es bueno para el aprendizaje del alumno y que es un buen sistema de evaluación para el profesor y para el estudiante. Esto se expone en las tablas 4.31, correspondiente a los ítems del *test por aciertos*, 4.32, correspondiente a los ítems del TBC y 4.33, correspondiente al examen de *preguntas abiertas*.

TABLA 4.31.

Correlaciones de los tres ítems que miden la *validez* del *test por aciertos*.

	El test por aciertos le permite al profesor conocer mejor... (ítem 15)	El test por aciertos me permite detectar mis lagunas formativas (ítem 18)
El test por aciertos me hace aprender más (ítem 21)	0.558	0.509

Nota. *r* de Pearson; significativa con una $p < 0.0001$

TABLA 4.32.

Correlaciones de los tres ítems que miden la *validez* del TBC.

	El TBC le permite al profesor conocer mejor... (ítem 16)	El TBC me permite detectar mis lagunas formativas (ítem 22)
El TBC me hace aprender más (ítem 22)	0.605	0.441

Nota. *r* de Pearson; significativa con una $p < 0.0001$

TABLA 4.33.

Correlaciones de los tres ítems que miden la *validez* del examen de *preguntas abiertas*.

	El examen de preguntas abiertas le permite al profesor conocer mejor... (17)	El examen de preguntas abiertas me permite detectar mis lagunas formativas (23)
El examen de preguntas abiertas me hace aprender más (ítem 23)	0.685	0.591

Nota. *r* de Pearson; significativa con una $p < 0.0001$

Hay, como se ve en las tablas 4.31, 4.32 y 4.33, una estrecha correlación entre creer que un tipo concreto de examen es para el profesor un buen sistema de evaluación (le proporciona una ajustada información de lo que el alumno sabe), que es para el alumno un buen procedimiento de autoevaluación (le permite conocer mejor sus lagunas formativas) y, en función de ello, que es un procedimiento de examen que hace que el alumno aprenda más. Los tres agrupamientos son consistentes y miden un aspecto que vamos a llamar «*validez percibida por el alumno*» del examen.

Sumando las puntuaciones obtenidas en cada pregunta por cada examen obtenemos el siguiente cuadro, que exponemos en la tabla 4.34.

TABLA 4.34.

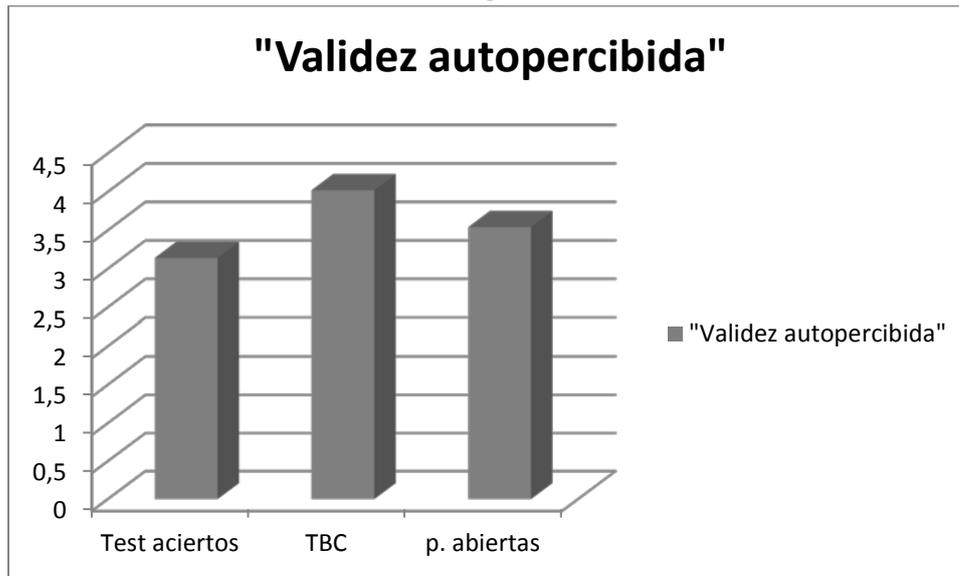
Media de las puntuaciones del agrupamiento *validez autopercebida* de cada examen.

Ítems de «validez» de los exámenes	Puntuaciones (de 3 a 18)	Puntuaciones medias de cada ítem (de 1 a 6)	Media sobre la puntuación máxima
Test <i>por aciertos</i> (15+18 +21)	12.41	4.14	62.6 % (ds 9.36)
TBC (16+19+22)	15.05	5.32	80.3 % (ds 7.69)
<i>Preguntas abiertas</i> (17+20+23)	13.61	4.54	70.7 % (ds 11.81)

Nota. En negrita el resultado más elevado de los tres tipos de examen. Las diferencias entre las medias de los tres tipos de examen son significativas a nivel de $p < 0.001$.

Las puntuaciones medias de cada examen en «validez autopercebida», expuestas en la tabla 4.34, las ponemos gráficamente en la figura 4.5.

Figura 4.5.



Según los datos de la tabla superior, podemos afirmar que los alumnos consideran que el TBC es el sistema de examen más válido.

En resumen, el índice de consistencia interna o fiabilidad (alfa de Cronbach) del agrupamiento de estos tres ítems es de 0.715 para el TBC, de 0.756 para el *test por aciertos* y de 0.809 para el examen de *preguntas abiertas*. Que los ítems que valoran la opinión positiva sobre este último tipo de examen tengan la mayor consistencia interna parece indicar que los partidarios del examen de *preguntas abiertas* constituyen un grupo muy definido de estudiantes.

- **Estudio de las correlaciones de los agrupamientos que miden la opinión del estudiante sobre la *validez autopercebida* de los exámenes.**

En primer lugar, veremos la correlación de los agrupamientos que miden la opinión sobre la *validez autopercebida* de los exámenes con los enfoques de aprendizaje y la autorregulación. El análisis de las correlaciones de los enfoques de aprendizaje y de la autorregulación con los tres agrupamientos que expresan la opinión positiva del alumno sobre cada tipo de examen arroja la tabla 4.35, extractada del anexo G:

TABLA 4.35.

Correlaciones del agrupamiento que expresa la «validez autopercebida» del examen.

	TBC 16+19+22	Test por aciertos 15+17+21	Preguntas a desarrollar 17+20+23
Ítems de <i>ENFOQUE PROFUNDO</i> (10 ítems)	0.298 p <0.0001	NS	0.127 p <0.05
Ítems de <i>MOTIVO PROFUNDO</i> (5 ítems)	0.307 p <0.0001	NS	0.136 p <0.05
Ítems de <i>ESTRATEGIA PROFUNDA</i> (5 ítems)	0.229 p <0.001	NS	NS
Ítems de <i>ENFOQUE SUPERFICIAL</i> (5 ítems)	-0.198 p <0.01	NS	NS
Ítems de <i>MOTIVO SUPERFICIAL</i> (5 ítems)	-0.245 p <0.001	NS	NS
Ítems de <i>ESTRATEGIA SUPERFICIAL</i> (5 ítems)	-0.119 p <0.05	-0.125 p <0.05	NS
Ítems de AUTORREGULACIÓN	0.195 p <0.01	NS	NS

Nota. Solo aparecen las correlaciones que son significativas (r de Pearson). Se destacan en negrita las correlaciones más altas en cada fila.

Como se ve, mientras que la buena opinión sobre el TBC se correlaciona de manera importante con los enfoques de aprendizaje profundos, especialmente con el *motivo profundo*, y de manera negativa con los superficiales, la buena opinión sobre los otros dos sistemas de examen no tiene apenas correlaciones con los enfoques, solo el considerar al examen de preguntas abiertas un buen sistema se correlaciona, pero muy débilmente con el *motivo profundo*.

Se puede afirmar, por lo tanto, que los alumnos con enfoques de aprendizaje profundos tienden a tener una buena opinión del TBC; lo contrario sucede con los que manifiestan enfoques superficiales. Esto no ocurre con los otros dos sistemas de examen.

En los ítems de autorregulación estudiados, también los alumnos con mejores puntuaciones en este aspecto tienden a tener una buena opinión del TBC, si bien aquí la correlación es algo más débil.

Esta correlación entre tener enfoques de aprendizaje profundo y pensar que el TBC es el mejor sistema de examen es paralela a la que existe con aquellos a los que les gusta el TBC como examen -ítem 3- y también con aquellos que opinan que el TBC es el sistema que les obliga a comprender más en el estudio, como hemos visto más arriba. Integrando todos estos ítems, que miden la preferencia de los alumnos por cada sistema de examen, su opinión sobre cuál es el que obliga a un mayor esfuerzo de comprensión y el agrupamiento de tres ítems que mide su opinión sobre la *calidad* -«validez autopercebida»- de cada examen, obtenemos la tabla 4.36, que para dividimos en tres tablas parciales para una más fácil visualización.

TABLA 4.36. A

Correlaciones de los ítems de satisfacción con los exámenes con los enfoques de aprendizaje y la autorregulación

ÍTEMS Y SU ENUCIADO ABREVIADO	Ítem 3 – «Me gusta el TBC»	Ítem 2 – «Me gusta el test por aciertos»	Ítem 4 – «Me gustan las preguntas abiertas»
Ítems de <i>ENFOQUE PROFUNDO</i>	0.225 p<0.001	NS	NS
Ítems de <i>MOTIVO PROFUNDO</i>	0.190 p<0.01	NS	NS
Ítems de <i>ESTRATEGIA PROFUNDA</i>	0.215 p<0.001	NS	NS
Ítems de <i>ENFOQUE SUPERFICIAL</i>	-0.142 p<0.05	NS	NS
Ítems de <i>MOTIVO SUPERFICIAL</i>	NS	NS	NS
Ítems de <i>ESTRATEGIA SUPERFICIAL</i>	-0.168 p<0.01	NS	NS
Ítems de <i>AUTORREGULACIÓN</i>	0.134 p<0.05	NS	NS

Nota. Se destacan en negrita las correlaciones con significación estadística de p<0.01 o menos.

TABLA 4.36. B

Correlaciones de los ítems de «esfuerzo de comprensión» con los enfoques de aprendizaje y la autorregulación

ÍTEMS Y SU ENUCIADO ABREVIADO	Ítem 9 – «el TBC me obliga a entender ...»	Ítem 8 – «el test por aciertos me obliga a entender...»	Ítem 10 – «El examen de preguntas abiertas me obliga a entender...»
Ítems de <i>ENFOQUE PROFUNDO</i>	0.303 p<0.0001	0.137 p<0.05	0.152 p<0.05
Ítems de <i>MOTIVO PROFUNDO</i>	0.316 p<0.0001	0.134 p<0.05	0.137 p<0.05
Ítems de <i>ESTRATEGIA PROFUNDA</i>	0.229 p<0.001	NS	0.136 p<0.05
Ítems de <i>ENFOQUE SUPERFICIAL</i>	-0.192 p<0.01	NS	-0.155 p<0.05
Ítems de <i>MOTIVO SUPERFICIAL</i>	-0.231 p<0.001	-0.163 p<0.01	-0.129 p<0.05
Ítems de <i>ESTRATEGIA SUPERFICIAL</i>	-0.168 p<0.01	NS	NS
Ítems de <i>AUTORREGULACIÓN</i>	0.134 p<0.05	NS	NS

TABLA 4.36. C

Correlaciones de los agrupamientos de *validez autopercebida* con los enfoques de aprendizaje y la autorregulación

	TBC 16+19+22	Test por aciertos 15+17+21	Preguntas a desarrollar 17+20+23
Ítems de ENFOQUE PROFUNDO	0.298 p<0.0001	NS	0.127 p<0.05
Ítems de MOTIVO PROFUNDO	0.307 p<0.0001	NS	0.136 p<0.05
Ítems de ESTRATEGIA PROFUNDA	0.229 p<0.001	NS	NS
Ítems de ENFOQUE SUPERFICIAL	-0.198 p<0.01	NS	NS
Ítems de MOTIVO SUPERFICIAL	-0.245 p<0.001	NS	NS
Ítems de ESTRATEGIA SUPERFICIAL	-0.119 p<0.05	-0.125 p<0.05	NS
Ítems de AUTORREGULACIÓN	0.195 p<0.01	NS	NS

Se ve el paralelismo entre los tres cuadros de correlaciones. Hay un detalle significativo, mientras que cuando el alumno responde que le gusta el TBC, esa respuesta positiva se relaciona más con las ESTRATEGIAS que con los MOTIVOS profundos, cuando lo que el alumno valora es si el TBC es un buen sistema de examen, con independencia de su gusto, la opinión positiva se relaciona más con los MOTIVOS que con las ESTRATEGIAS; y esto ocurre tanto con las correlaciones positivas con los enfoques profundos como con las negativas con los superficiales.

En la tabla 4.37 exponemos las correlaciones de la opinión del alumno sobre la validez de los diversos tipos de examen y los demás ítems de la encuesta. Para mayor

claridad, ponemos solo aquellos ítems o agrupamientos de ítems que tienen alguna r superior a 0,20, correlación significativa a un $p < 0.001$.

TABLA 4.37.

Otras correlaciones de los agrupamientos de «validez autopercebida» de los exámenes.

	TBC 16+19+22	Test por aciertos 15+17+21	Preguntas a desarrollar 17+20+23
ítem 2 – Examen que <i>más le gusta</i> : test por aciertos	0.245 $p < 0.001$	NS	NS
ítem 3 – Examen que <i>más le gusta</i> : TBC	0.309 $p < 0.0001$	-0.175 $p < 0.01$	NS
ítem 4 – Examen que <i>más le gusta</i> : preguntas abiertas	NS	NS	0.288 $p < 0.0001$
ítem 5 – Examen que te <i>hace estudiar más</i> : test por aciertos	0.193 $p < 0.01$	0.383 $p < 0.0001$	0.215 $p < 0.001$
ítem 6 – Examen que te <i>hace estudiar más</i> : TBC	0.382 $p < 0.0001$	0.121 $p < 0.05$	0.243 $p < 0.001$
ítem 7 – Examen que te <i>hace estudiar más</i> : preguntas abiertas	0.168 $p < 0.01$	0.202 $p < 0.001$	0.266 $p < 0.001$
Suma de ítems 5+6 y 7 – Dedicar mucho tiempo al estudio con todos los sistemas de examen	0.327 $p < 0.0001$	0.326 $p < 0.0001$	0.326 $p < 0.0001$
ítem 8 – Examen que obliga a <i>comprender</i> : test por aciertos	0.157 $p < 0.05$	0.550 $p < 0.0001$	NS
ítem 9 – Examen que obliga a <i>comprender</i> : El TBC	0.415 $p < 0.0001$	0.201 $p < 0.01$	NS
ítem 10 – Examen que obliga a <i>comprender</i> : preguntas abiertas	NS	NS	0.647 $p < 0.0001$
ítem 11 – Examen que obliga a <i>memorizar</i> : test por aciertos	NS	0.182 $p < 0.01$	0.266 $p < 0.001$
ítem 12 – Examen que obliga a <i>memorizar</i> – Ítem 12: El TBC	0.251 $p < 0.001$	NS	0.366 $p < 0.0001$
ítem 13 – Examen que obliga a <i>memorizar</i> – Ítem 13: preguntas abiertas	0.182 $p < 0.01$	0.225 $p < 0.001$	NS
ítem 24 – En el examen que hice, <i>escogí bien mi grado de seguridad</i> (ítem 24)	0.211 $p < 0.001$	NS	NS
Suma de ítems 25 – <i>Me estresa mucho hacer un examen de tipo test por aciertos</i>	NS	0.258 $p < 0.001$	NS
Suma de ítems 15+18+21 – Validez del test por aciertos	0.317 $p < 0.0001$	1	NS $p < 0.001$
Suma de ítems 16+19+22 – Validez del TBC	1	0.317 $p < 0.0001$	NS
Suma de ítems 17+20+23 – Validez del examen de preguntas abiertas.	NS	NS	1

Nota. Se resaltan en negrita las correlaciones elevadas, superiores a 0.40, r de Pearson.

Hemos resaltado en negrita las correlaciones superiores a 0.40; hay tres correlaciones superiores a esta cifra y corresponden a la pregunta sobre qué examen obliga a entender mejor lo que se estudia. Con esa pregunta, la correlación es importante con aquel examen que el alumno considera mejor (el TBC con una r de 0.415, el test por aciertos con una r de 0.550 y el examen de preguntas abiertas con una r de 0.647). Estas altas correlaciones indican que en la opinión positiva que un alumno tiene de un tipo de examen, el considerar que te obliga a comprender mejor lo que estudias es un factor muy importante. La *validez percibida por el alumno* de un tipo de examen está muy relacionada con la idea de que dicho examen obligue a entender mejor los conceptos que se estudian. En la tabla 4.38 exponemos las correlaciones de los tres ítems que valoran la comprensión con el agrupamiento *validez autopercebida*:

TABLA 4.38.

Correlación con ítems que valoran la necesidad de comprender bien los conceptos estudiados

	TBC 16+19+22	Test por aciertos 15+17+21	Preguntas a desarrollar 17+20+23
ítem 8 – <i>Examen que obliga a comprender: test por aciertos</i>	0.157 $p < 0.05$	0.550 $p < 0.0001$	NS
ítem 9 – <i>Examen que obliga a comprender: El TBC</i>	0.415 $p < 0.0001$	0.201 $p < 0.01$	NS
ítem 10 – <i>Examen que obliga a comprender</i> ¹⁰ : <i>preguntas abiertas</i>	NS	NS	0.647 $p < 0.0001$

Nota. Se resaltan en negrita las correlaciones elevadas, superiores a 0.40, r de Pearson.

En la opinión de los alumnos hay una correlación importante entre considerar que un tipo de examen obliga a comprender mejor lo que se estudia y considerarlo un buen sistema de examen.

De las demás correlaciones extraemos como conclusión que:

– Tipos de exámenes que a los alumnos les gusta más:

El examen que a los alumnos **más les gusta** (ítems 2, 3 y 4) tiende a ser el que consideran mejor. Si les gusta más el TBC, tienden ligeramente a tener peor opinión del test por aciertos (correlación resaltada en negrita), como se ve en la tabla 4.39.

TABLA 4.39.

Correlación del agrupamiento que valora la «validez» de cada examen con ítems que valoran el aprecio del alumno con cada tipo de examen

	Validez TBC	Validez test por aciertos	Validez preguntas abiertas
ítem 2 – Examen que más le gusta : test por aciertos	NS	0.245 (p<0.001)	NS
ítem 3 – Examen que más le gusta : TBC	0.309 (p<0.0001)	-0.175 (p<0.01)	NS
ítem 4 – Examen que más le gusta : preguntas abiertas	NS	NS	0.288 (p<0.0001)

– Examen que obliga al alumno a dedicar más tiempo a estudiar:

La tabla de correlaciones del agrupamiento que valora la *validez* de los exámenes con el *tiempo* que el alumno ha de dedicar a prepararlo muestra una correlación clara entre ambos conceptos, «*un buen sistema de examen obliga a dedicar mucho tiempo al estudio*». Como podemos ver en la tabla 4.40, todas las correlaciones son significativas, aunque desigualmente altas. La tendencia a puntuar alta esta pregunta de dedicación de tiempo a la preparación de los exámenes en todos los ítems (suma de 5 + 6 + 7), que de alguna manera expresa una apreciación subjetiva de dedicar mucho tiempo al estudio, sea cual sea el tipo de examen, se correlaciona bien con la opinión positiva sobre todos

los sistemas de examen, no solo con uno. Los alumnos que dicen dedicar más tiempo al estudio tienen buena opinión de todos los tipos de examen:

TABLA 4.40.

Correlaciones con los ítems que valoran el tiempo dedicado al estudio con cada examen con los agrupamientos.

	<i>Validez Autopercebida del TBC</i> 16+19+22	<i>Validez A. del test por aciertos</i> 15+17+21	<i>Validez A. de las preguntas a desarrollar</i> 17+20+23
Ítem 5- Examen que te <i>hace estudiar más</i> : test por aciertos	0.193 (p<0.01)	0.382 (p<0.0001)	0.215 (p<0.001)
ítem 6- Examen que te <i>hace estudiar más</i> : TBC	0.382 (p<0.0001)	0.121 (p<0.05)	0.243 (p<0.001)
ítem 7- Examen que te <i>hace estudiar más</i> : preguntas abiertas	0.168 (p<0.01)	0.202 (p<0.001)	0.266 (p<0.001)
Dedicar mucho tiempo al estudio con todos los sistemas de examen (5 + 6 + 7)	0.327 (p<0.0001)	0.326 (p<0.0001)	0.326 (p<0.0001)

Nota. Resaltadas en negrita las correlaciones mayores de cada ítem. Extraído del anexo G.

Como se ve en la tabla 4.40, hay correlaciones significativas de cualquier ítem que mide la cantidad de estudio con cualquier agrupamiento que valora la opinión positiva sobre los exámenes: **los alumnos reconocen que obligar a estudiar más es una característica propia de un buen sistema de evaluación.** Un dato interesante de esta tabla es que aparece una correlación entre dedicar mucho tiempo a la preparación de los exámenes, con cualquier sistema, y tener una opinión positiva de cualquier sistema de examen, tal como se ve en la tabla 4.37 en las correlaciones cruzadas, todas positivas y significativas. La opinión negativa sobre los tres sistemas de evaluación, que podríamos considerar una opinión negativa sobre los exámenes en general, se asocia a la sensación de dedicar poco tiempo al estudio; y lo contrario sucede con la opinión positiva. El alumno que, en su opinión, dedica poco tiempo a la preparación de los

exámenes es más crítico frente a todos; el que reconoce estudiar mucho tiene mejor opinión de todos ellos.

– **Exámenes que obligan al alumno a memorizar más datos:**

La tabla 4.41 muestra las correlaciones entre el agrupamiento de ítems que juzga la que hemos llamado *validez percibida* de un tipo de examen y la apreciación del esfuerzo de memorización que los exámenes.

TABLA 4.41.

Correlaciones con los ítems que valoran esfuerzo de memorización con cada examen con los agrupamientos; extraído del anexo G.

	<i>Validez autopercebida del TBC</i>	<i>Validez A. del test por aciertos</i>	<i>Validez A. de las preguntas a desarrollar</i>
Ítem 11- Examen que obliga a memorizar : test por aciertos	NS	0.182 (p<0.01)	0.266 (p<0.001)
Ítem 12- Examen que obliga a memorizar : TBC	0.251 (p<0.001)	NS	0.366 (p<0.0001)
Ítem 13- Examen que obliga a memorizar : examen de preguntas abiertas	0.182 (p<0.01)	0.225 (p<0.001)	NS

Nota. En negrita las correlaciones superiores a 0.20

La correlación de la pregunta sobre el esfuerzo de memorización a que obliga cada examen con la opinión positiva sobre los tipos de examen arroja resultados diferentes a los otros ítems que hemos visto anteriormente. En general, los alumnos que opinan que los exámenes de tipo test son exámenes de alta *validez* suelen opinar que el examen de preguntas abiertas les obliga a memorizar mucho, mientras que aquellos que consideran a las preguntas abiertas un buen sistema de examen creen que los de tipo test, especialmente el test por aciertos, también les obligan a memorizar mucho. Los que defienden la validez de las preguntas abiertas como sistema de evaluación no asocian

esta opinión a un mayor tiempo de dedicación al estudio. Esta tabla hay que ponerla en relación con la que expresaba el tiempo que se le dedica al estudio con cada tipo de examen, que era:

TABLA 4.42.

Tabla que muestra el esfuerzo de memorización que exige cada examen

	TBC (Ítem 12)	Test <i>por aciertos</i> (Ítem 11)	<i>Preguntas abiertas</i> (Ítem 13)
Esfuerzo de memorización que el examen exige (1 a 6)	4.82	4.27	5.27

Nota. Todas las diferencias entre los tres ítems son significativas, r de 0.001. $F = 68.04$, $p < 0.0001$, $\eta^2 = 0.201$.

Dentro del dato que se expresa en esa tabla, de que el examen de *preguntas abiertas* es el que obliga al alumno a una mayor memorización y el test por aciertos el que menos obliga a *memorizar*, los alumnos que tienen buena opinión de la validez de los test (en los dos sistemas, TBC y *por aciertos*) tienden a aumentar su estimación del esfuerzo de memorización que el examen de *preguntas abiertas* exige; en sentido inverso, los alumnos que tienen la opinión de que el examen de preguntas abiertas es un buen sistema de examen tienden a opinar que los test obligan a memorizar mucho.

– Haber escogido con acierto el grado de seguridad en el TBC (ítem 24):

Ante el ítem que valora la opinión del alumno sobre el acierto que ha tenido al escoger su grado de seguridad en el examen, el ítem 24 *-En el examen que hice escogí bien mi grado de seguridad-*, la correlación con la *validez percibida* de cada tipo de examen se muestra en la tabla 4.43.

TABLA 4.43.

Correlaciones de la «validez autopercebida» -V.A.- con el acierto al escoger la opción de seguridad.

	V.A. del TBC 16+19+22	V.A. del test por aciertos 15+17+21	V.A. de las preguntas a desarrollar 17+20+23
<i>En el examen que hice, escogí bien mi grado de seguridad</i> (ítem 24)	0.211 (p<0.001)	NS	NS

Hay una correlación significativa entre haber tenido un buen desempeño con el TBC, en el sentido de haber escogido bien la opción de seguridad, y tener la opinión de que el TBC es un buen sistema de examen (r de 0.211). Este hallazgo parece natural, pero si comparamos la magnitud de esa correlación, más bien ligera, con la correlación que hay entre haberse desempeñado bien con el examen y tener preferencia, o gusto, por el tipo de examen, encontramos que en el segundo caso la correlación es mucho más alta, con una r de 0.534, tal como se ve en la tabla 4.44:

TABLA 4.44.

Correlaciones entre los exámenes preferidos por los alumnos y su opinión sobre la «validez autopercebida» de cada tipo de examen; extraído del anexo G

ÍTEMS DE «EL EXAMEN QUE MÁS GUSTA »	Me gusta el TBC (3)	Me gusta el test por aciertos (2)	Me gustan las preguntas a desarrollar(4)
<i>En el examen que hice (TBC) escogí bien mi grado de seguridad</i> (ítem 24)	0.534 (p<0.0001)	-0.131 (p<0.05)	NS

La comparación de estas dos correlaciones indica que, en lo tocante a la preferencia del alumno por el TBC, influye más la experiencia positiva con él -en el sentido de haber estado afortunado escogiendo la opción de seguridad- que la opinión de que el TBC sea un sistema de evaluación objetivamente bueno. En cualquier caso, hay correlación con ambos conceptos. La gran diferencia de ambas correlaciones, 0.53 frente a 0.21, parece indicar que incluso los alumnos que reconocen que no les ha ido

bien escogiendo su grado de seguridad tienen, a pesar de eso, buena opinión sobre la validez del TBC.

Análisis de la correlación de la *validez autopercebida* de cada examen con el estrés inducido al alumno

En la tabla 4.45 podemos observar las correlaciones entre la alta puntuación en *estrés con un examen* y las opiniones del alumno sobre la *validez* de cada examen.

TABLA 4.45.

Correlación entre el nivel de estrés que genera cada examen en el alumno y su opinión sobre la «validez autopercebida» (VA) de cada examen.

Opinión del alumno sobre la validez del examen	V.A. del TBC 16+19+22	V. A. del test por aciertos 15+17+21	V. A. de las preguntas a desarrollar 17+20+23
Ítem 26 - Me estresa mucho hacer un examen de tipo <i>test por aciertos</i>	NS	0.258 ($p < 0.001$)	NS
Ítem 27 - Me estresa mucho hacer un examen de tipo TBC	NS	0.156 ($p < 0.05$)	NS
Ítem 28 - Me estresa mucho hacer un examen de <i>preguntas abiertas</i>	NS	NS	NS

Como se puede ver en la tabla 4.45, no hay correlación entre el estrés que le induce un examen al alumno y la opinión sobre la *validez percibida por el alumno* de los distintos sistemas de examen, salvo en el caso de los alumnos que tienen buena opinión del test por aciertos, que tienden a mostrar mayor estrés con ese tipo de examen y, en menor medida, con el otro sistema de test (TBC). En este caso, da la sensación de que los estudiantes que opinan que el test tradicional es un buen sistema de evaluación tienden a considerarlo un sistema de examen más exigente, que les obliga a estudiar mucho (r de 0.38), a comprender bien lo que aprenden (r de 0.55), a memorizar mucho

(r de 0.18) y que les genera estrés (r de 0.26). Con los otros dos sistemas de examen analizados, el TBC y el examen de preguntas abiertas, vemos las correlaciones de la opinión elevada sobre la *validez percibida* de cada sistema evaluatorio con el tiempo de estudio y el esfuerzo de comprensión que exigen, pero no hay correlación con el nivel de estrés.

Finalmente, correlacionando la opinión positiva de cada tipo de examen con los demás, obtenemos la tabla 4.46.

TABLA 4.46.

Correlaciones de los grupos que valoran la opinión del alumno sobre la «validez autopercibida» de los exámenes.

	El test <i>por aciertos</i> es válido 15+17+21	El TBC es válido 16+19+22	las <i>preguntas a desarrollar</i> son válidas 17+20+23
El test <i>por aciertos</i> es válido (VA) 15+17+21	1		
El TBC es <i>válido</i> (VA) 16+19+22	0.317 ($p < 0.0001$)	1	
El examen <i>preguntas a desarrollar</i> es válido (VA) 17+20+23	NS	NS	1

En la tabla vemos que los partidarios, por considerarlos de alta validez, de cualquiera de los exámenes de tipo test, sea *por aciertos* o por TBC, tienden a tener una buena opinión del otro tipo de test (r de 0.32). En cambio, no hay correlación, ni negativa ni positiva, entre tener una buena opinión del examen por preguntas abiertas y tenerla de los exámenes de tipo test.

**4.2.1.5. AGRUPAMIENTO DE LAS PREGUNTAS QUE MIDEN LA «EFICACIA DOCENTE»
PERCIBIDA POR EL ALUMNO DE CADA TIPO DE EXAMEN**

Hay otras dos preguntas que tienen una alta correlación con el agrupamiento anterior, que medía la opinión del alumno sobre la validez percibida por el alumno de cada tipo de examen. Son las siguientes:

– *«¿En qué medida cada tipo de examen te obliga a dedicar más tiempo al estudio?»*

Ítem 5) *Test tradicional, sin indicar grado de seguridad.*

Ítem 6) *Test teniendo que indicar mi grado de seguridad.*

Ítem 7) *Examen de tipo desarrollar un tema.*

– *«¿En qué medida cada tipo de examen te obliga a entender bien todo lo que estás estudiando?»*

Ítem 8) *Test tradicional, sin indicar grado de seguridad.*

Ítem 9) *Test teniendo que indicar mi grado de seguridad*

Ítem 10) *Examen de tipo desarrollar un tema*

La tabla 4.47 muestra las correlaciones de los seis ítems anteriores con el agrupamiento que denominamos «validez autopercebida» de cada examen.

TABLA 4.47.

Correlación entre la «V.A.» de cada tipo de examen y el tiempo dedicado al estudio y el esfuerzo de comprensión

	VA del TBC 16+19+22	VA del test por aciertos 15+17+21	VA de las preguntas a desarrollar 17+20+23
Ítem 5 – Examen que te <i>hace dedicar más tiempo al estudio</i> : test por aciertos	0.193 (p<0.01)	0.382 (p<0.0001)	0.215 (p<0.001)
Ítem 6 – Examen que te <i>hace dedicar más tiempo al estudio</i> : TBC	0.382 (p<0.0001)	0.121 (p<0.05)	0.243 (p<0.001)
Ítem 7 – Examen que te <i>hace dedicar más tiempo al estudio</i> : preguntas abiertas	0.168 (p<0.01)	0.202 (p<0.001)	0.266 (p<0.001)
Ítem 8 – Examen que obliga a <i>comprender</i> : test por aciertos	0.157 (p<0.05)	0.550 (p<0.0001)	NS
Ítem 9 – Examen que obliga a <i>comprender</i> : TBC	0.415 (p<0.0001)	0.201 (p<0.001)	NS
Ítem 10 – Examen que obliga a <i>comprender</i> : preguntas abiertas	NS	NS	0.647 (p<0.0001)

Nota. Se resaltan en negrita las correlaciones superiores a 0.30

Estas dos preguntas, que antes estudiamos mediante sus estadísticos básicos, miden la opinión de los alumnos sobre el esfuerzo al que les obliga cada tipo de examen en dos aspectos: el tiempo dedicado al estudio y el esfuerzo para entender bien lo que se aprende. Uniéndolas al agrupamiento de las tres preguntas anteriores se obtiene un nuevo agrupamiento de cinco preguntas que miden la opinión del alumno sobre qué tipo de examen les hace trabajar más y es más válido. Puesto que hacer estudiar más, obligar a comprender mejor lo que se estudia y ser un procedimiento válido de evaluación definen todos los rasgos que hacen de un sistema de evaluación un procedimiento eficaz de aprendizaje, vamos a denominar a este agrupamiento de ítems *eficacia docente percibida por el alumno* de cada tipo de examen (a partir de ahora *eficacia docente*, ara abreviar):

- Le obliga a **estudiar más** (ítems 5, 6 y 7, para cada tipo de examen)

- Le obliga a **entender mejor** la materia (8, 9 y 10)
- Permite que **el profesor sepa mejor** lo que el alumno sabe (15, 16 y 17)
- Hace que **el alumno se dé cuenta** mejor de sus **lagunas** formativas (18, 19 y 20)
- Hace que el alumno **aprenda más** (21, 22 y 23)

Aparecen, por lo tanto, tres agrupamientos de cinco elementos, referidos a cada tipo de examen:

TABLA 4.48.

Agrupamientos que definen la *eficacia docente* de cada tipo de examen.

Opinión del alumno sobre la «eficacia docente» del <i>test basado en aciertos</i>	Ítems 5, 8, 15, 18 y 21
Opinión del alumno sobre la «eficacia docente» del TBC	Ítems 6, 9, 16, 19 y 22
Opinión del alumno sobre la «eficacia docente» del examen de <i>preguntas abiertas</i>	Ítems 7, 10, 17, 20 y 23

Estos agrupamientos expresarían el siguiente concepto: «*preguntas que piden la opinión del alumno sobre si un tipo de examen es adecuado para evaluarle y le hace trabajar más, comprender mejor, autoevaluarse y aprender más*».

Las correlaciones internas (resumidas del anexo G) y la fiabilidad de cada uno de estos agrupamientos de *eficacia docente* se muestran en las tablas que exponemos a continuación.

- Tablas de correlaciones del factor *eficacia docente* con cada examen.

Para el *test por aciertos*, en la tabla 4.49 exponemos, en primer lugar, las correlaciones internas del agrupamiento.

TABLA 4.49.

Correlaciones internas de los ítems de *eficacia docente* del test *por aciertos*.

Correlaciones	5	8	15	18	21
5	1				
8	0.370	1			
15	0.352	0.451	1		
18	0.233	0.385	0.463	1	
21	0.368	0.514	0.558	0.509	1

Nota. *r* de Pearson. Todas son significativas con $p < 0.001$

La tabla de correlaciones de cada ítem con el grupo (*peso* de cada ítem) se muestra en la tabla 4.50.

TABLA 4.50.

Eficacia docente del test *por aciertos*: tabla de correlaciones de cada ítem con el agrupamiento (*peso* de cada ítem)

ítem	Peso del ítem	Enunciado del ítem
8	0.807	<i>El test por aciertos abiertas obliga a comprender lo que se estudia</i>
21	0.785	<i>El test por aciertos me está haciendo aprender más</i>
15	0.702	<i>El test por aciertos le hace al profesor tener mejor información de mi aprendizaje</i>
18	0.552	<i>El test por aciertos me hace darme cuenta de mis lagunas formativas</i>
5	0.450	<i>El test por aciertos hace dedicar más tiempo al estudio</i>

Nota. El *peso del ítem* es la correlación de cada ítem con la suma de todos menos el propio ítem, obtenidas mediante el SPSS (*r* de Pearson; análisis detallado en el capítulo 4.4.4. y en anexo H). $p < 0.0001$

El índice α de Cronbach de este agrupamiento, ítems 5, 8, 15, 18 y 21, es de **0.785**.

Para el **TBC**, en la tabla 4.51 exponemos, en primer lugar, las correlaciones internas de los ítems del agrupamiento y en la 4.49 las de cada ítem con el agrupamiento.

TABLA 4.51.

Correlaciones internas de los ítems de *eficacia docente* del TBC.

r de Pearson	6	9	16	19	22
6	1				
9	0.282	1			
16	0.314	0.322	1		
19	0.322	0.270	0.346	1	
22	0.283	0.369	0.605	0.441	1

Nota. r de Pearson. Todas significativas $p < 0.001$.

La tabla de correlaciones de cada ítem con el agrupamiento (peso de cada ítem) se muestra en la tabla 4.52.

TABLA 4.52

Eficacia docente del TBC. Tabla de correlaciones de cada ítem con los demás del agrupamiento o *peso* cada ítem

ítem	Peso del ítem	Enunciado del ítem
9	0.717	<i>El TBC me obliga a comprender lo que se estudia</i>
16	0.696	<i>El TBC le hace al profesor tener mejor información de mi aprendizaje</i>
22	0.665	<i>El TBC me está haciendo aprender más</i>
6	0.618	<i>El TBC hace dedicar más tiempo al estudio</i>
19	0.451	<i>El TBC me hace darme cuenta de mis lagunas formativas</i>

Nota. El *peso del ítem* es la correlación de cada ítem con la suma de todos menos el propio ítem, obtenidas mediante el SPSS (r de Pearson; análisis detallado en el capítulo 4.4.4. y en anexo H). $p < 0.0001$

El índice α de Cronbach del agrupamiento correspondiente al TBC (6, 9, 19, 19 y 22) es de **0.737**.

Este agrupamiento, de 5 ítems, en el caso del TBC, es el que da una mayor consistencia interna; si se le añade otro ítem, como el 12 («*el TBC me hace memorizar más*»), la consistencia disminuye; así mismo, el alfa de Cronbach también baja si se excluye cualquiera de sus 5 ítems.

Para el **examen de preguntas abiertas**, en la tabla 4.53 exponemos, en primer lugar, las correlaciones internas de los ítems del agrupamiento y en la 4.51 las de cada ítem con el agrupamiento.

TABLA 4.53.

Correlaciones internas de los ítems de *eficacia docente* del *examen de preguntas abiertas*.

r de Pearson	7	10	17	20	23
5	1				
10	0.318 p<0.0001	1			
17	0.142 p<0.05	0.562 p<0.0001	1		
20	0.345 p<0.0001	0.528 p<0.0001	0.480 p<0.0001	1	
23	0.197 p<0.01	0.563 p<0.0001	0.685 p<0.0001	0.591 p<0.0001	1

Nota. r de Pearson.

TABLA 4.54.

Eficacia docente del examen de preguntas abiertas: la tabla de correlaciones de cada ítem con el agrupamiento (peso del ítem).

ítem	Peso del ítem	Enunciado del ítem
20	0.804	<i>El examen de preguntas abiertas me hace darme cuenta de mis lagunas formativas</i>
23	0.795	<i>El examen de preguntas abiertas me está haciendo aprender más</i>
10	0.792	<i>El examen de preguntas abiertas obliga a comprender lo que se estudia</i>
17	0.759	<i>El examen de preguntas abiertas le hace la profesor tener mejor información de mi aprendizaje</i>
7	0.511	<i>El examen de preguntas abiertas hace dedicar más tiempo al estudio</i>

Nota. El *peso del ítem* es la correlación de cada ítem con la suma de todos menos el propio ítem, obtenidas mediante el SPSS (r de Pearson; análisis detallado en el capítulo 4.4.4. y en anexo H). p<0.0001

El índice α de Cronbach del agrupamiento correspondiente al examen de *preguntas abiertas* es de **0.891**, el más alto de los tres agrupamientos.

Es de destacar la elevada consistencia interna de este agrupamiento, la mayor de los tres, con un α de más de 0.89. En este caso se ve que el ítem 7, «*me obliga a dedicar más tiempo al estudio*», funciona algo peor que los demás, por lo que el agrupamiento de los otros cuatro ítems tendría mayor consistencia interna que el de los cinco, pero optamos por mantenerlo en nuestro análisis para poder compararlo con los dos agrupamientos anteriores. Esta elevada consistencia interna del agrupamiento del examen de preguntas abiertas parece indicar que el grupo de alumnos con una buena opinión sobre esta modalidad de examen es un grupo de características más homogéneas que los otros dos.

La correlación entre los valores de los tres agrupamientos, correspondientes al constructo que hemos llamado «*eficacia docente*», de los tres tipos de examen se muestra en la tabla 4.55.

TABLA 4.55.

Correlación entre los tres agrupamientos, correspondientes a cada tipo de examen, que miden la *eficacia docente*.

	Test por aciertos 5+8+15+18+21	TBC 6+9+16+19+22	Preguntas abiertas 7+10+17+20+23
Test por aciertos 5+8+15+18+21	1		
TBC 6+9+16+19+22	0.361 ($p < 0.0001$)	1	
Preguntas abiertas 7+10+17+20+23	0.021 (no significativo)	0.114 (no significativo)	1

Nota. r de Pearson.

Hay una correlación de cierta importancia -0.36- entre tener una buena opinión de la eficacia docente con cualquiera de los dos sistemas de test; es decir, los alumnos que tienen una buena opinión –en cuanto a eficacia docente- del test por aciertos la tienen también de TBC y viceversa. El examen de preguntas abiertas, en cambio, se comporta de forma independiente a los otros dos.

Como hemos visto, el análisis factorial con rotación ortogonal de componentes principales (rotación VARIMAX) confirma nuestros constructos, calculados a partir de la tabla de correlaciones (ver más adelante capítulo 4.4.4. y anexo H). En nuestra encuesta aparece un factor, que llamamos, eficacia docente de cada tipo de examen, con buena consistencia interna, muy buena en el caso del examen de preguntas abiertas, y que nos indica la opinión del alumno sobre la validez y el estímulo al aprendizaje – mayor tiempo de estudio, mayor esfuerzo de comprensión- de cada tipo de examen. Al haber preguntado sobre tres modalidades de examen, el test basado en aciertos, el TBC y el examen de preguntas abiertas, obtenemos tres componentes «*eficacia docente del examen*», uno para cada tipo. Esto nos permite poseer un instrumento consistente para analizar la opinión del alumno sobre la evaluación.

- **Puntuaciones que los estudiantes otorgan a cada examen en «*eficacia docente*»**

Aceptados, pues, los tres componentes que miden «*eficacia docente*» de cada examen, vamos a comparar los resultados de la encuesta en este factor con los tres exámenes comparados. Lo expresamos en la tabla 4.56.

TABLA 4.56.

Puntuación media en la encuesta a cada uno de los agrupamientos de *eficacia docente*.

<i>Eficacia docente de cada examen</i>	Puntuación media (1 a 6)	D.S.	Porcentaje sobre puntuación máxima
Test basado en el número de aciertos(5+8+15+18+21)	4.28 [4.24 - 4.32]	1.367	65.6 %
TBC (6+9+16+19+22)	5.09 [5.05-5.13]	1.047	81.8 %
Examen de preguntas abiertas (7+10+17+20+23)	4.73 [4.68 - 4.78]	1.650	74.6 %

Nota. Todas las diferencias son estadísticamente significativas con $p < 0.001$. Se resalta en negrita la media más elevada.

El TBC es superior a los otros dos sistemas de examen, con mayor diferencia en relación al test convencional (tamaño del efecto 0.59) y menor respecto al examen de preguntas abiertas (tamaño de efecto 0.22). Todas las diferencias son estadísticamente significativas con $p < 0.001$ (ANOVA), como se ve en la tabla 4.57

TABLA 4.57.

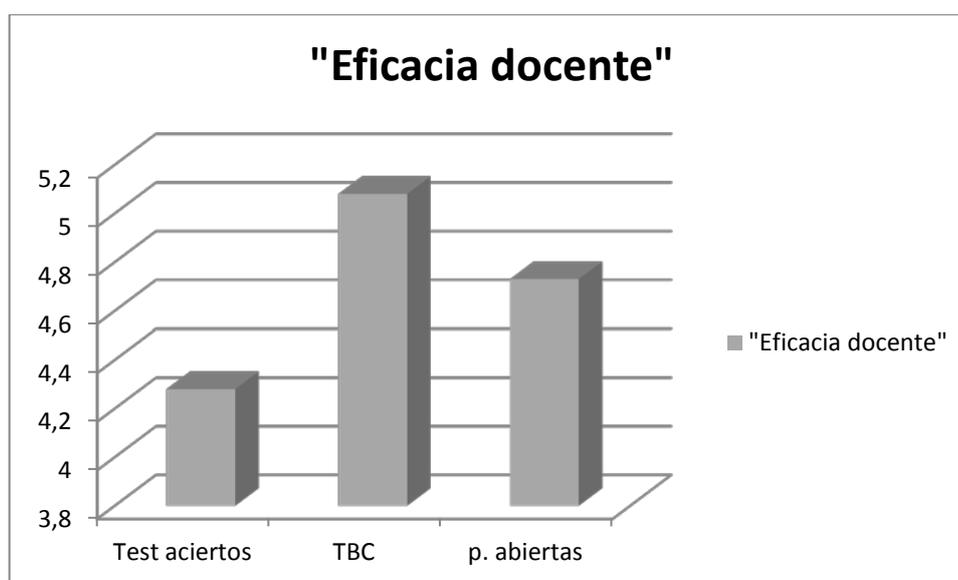
Significación estadística de las diferencias en la media con los tres exámenes.

	Test por aciertos	TBC
TBC	$p < 0.001$	–
Preguntas abiertas	$p < 0.001$	$p < 0.001$
$F = 77.99$	$p < 0.001$	$\text{Eta}^2 = 0.225$

Nota. Encuestas realizadas en 2013 y 2014; 272 estudiantes.

En la figura 4.6 podemos ver gráficamente estos valores de puntuación media asignada por los alumnos a cada agrupamiento.

Figura 4.6.



Los alumnos opinan que el TBC es el examen *más eficaz* para el aprendizaje del estudiante, en el sentido en que hemos definido la eficacia docente: es el más adecuado para evaluar al alumno, le hace trabajar más, le obliga a un mayor esfuerzo de comprensión y es el que induce un mayor aprendizaje.

– Correlaciones de la opinión sobre la eficacia docente de los sistemas de examen-agrupamientos de 5 ítems- y la encuesta sobre los enfoques de aprendizaje y la autorregulación.

La correlación que hemos obtenido entre estos tres agrupamientos y los enfoques de aprendizaje y la autorregulación se muestra en la tabla 4.58.

TABLA 4.58.

Correlación de *eficacia docente* de cada examen con enfoques de aprendizaje y autorregulación

	TBC 6+9+16+19+22	Test por aciertos 5+8+15+17+21	Preguntas a desarrollar 7+10+17+20+23
Ítems de ENFOQUE PROFUNDO	0.354 (p<0.0001)	NS	0.14 (p<0.05)
Ítems de MOTIVO PROFUNDO	0.360 (p<0.0001)	NS	0.15 (p<0.05)
Ítems de ESTRATEGIA PROFUNDA	0.278 (p<0.001)	NS	NS
Ítems de ENFOQUE SUPERFICIAL	-0.228 (p<0.001)	NS	NS
Ítems de MOTIVO SUPERFICIAL	-0.273 (p<0.001)	NS	NS
Ítems de ESTRATEGIA SUPERFICIAL	-0.15 (p<0.05)	-0.13 (p<0.05)	NS
Ítems de AUTORREGULACIÓN	0.233 (p<0.001)	0.12 (p<0.05)	0.13 (p<0.05)

Nota. Se resaltan las correlaciones mayores de 0.202, que tienen una p<0.001 (r de Pearson).

El grupo que valora más al TBC en cuanto a su *eficacia docente* se correlaciona con intensidad con los ítems de enfoque profundo de aprendizaje (0.35) y se correlaciona negativamente con los de enfoque superficial (-0.23). La correlación es más fuerte en ambos casos con los ítems de *motivo* (+0.36 con motivo profundo, -0.27 con motivo superficial) que con los de *estrategia* (+0.28 con *estrategia profunda*, -0.15 con *estrategia superficial*); con el agrupamiento de tres ítems que anteriormente llamamos «opinión positiva sobre el TBC» pasaba algo similar, aunque los niveles de alfa eran menores, la correlación era también mayor con los ítems de *motivo profundo* que con los de *estrategia profunda*.

Como se puede ver en la tabla, los agrupamientos correspondientes a los otros dos tipos de examen, *test por aciertos* y *preguntas a desarrollar*, no se correlacionan con

los enfoques de aprendizaje (lo que ocurre en la mayoría de los casos) o tienen una correlación muy pequeña, siempre menor que la correlación con el TBC.

Estos resultados se pueden comparar con los que obtuvimos en estos mismos ítems de enfoques de aprendizaje y autorregulación en el agrupamiento anterior de tres ítems –«*validez autopercebida*» de cada examen-. Se ve que las correlaciones con el TBC son mayores y que sigue sin haber correlaciones importantes con la opinión sobre los otros dos sistemas de examen.

A continuación exponemos, mediante la tabla 4.59, la comparación de los dos agrupamientos, de tres y de 5 ítems, referidos al TBC en relación con los enfoques de estudio y la autorregulación.

TABLA 4.59

Correlación de los agrupamientos *validez autopercebida* y *eficacia docente* de la opinión sobre el TBC con enfoques de aprendizaje y autorregulación

	V.A. del TBC 16+19+22	Eficacia del TBC 6+9+16+19+22
Ítems de ENFOQUE PROFUNDO	0.298 (p<0.0001)	0.354 (p<0.0001)
Ítems de MOTIVO PROFUNDO	0.307 (p<0.0001)	0.360 (p<0.0001)
Ítems de ESTRATEGIA PROFUNDA	0.229 (p<0.001)	0.278 (p<0.0001)
Ítems de ENFOQUE SUPERFICIAL	-0.198 (p<0.01)	-0.228 (p<0.001)
Ítems de MOTIVO SUPERFICIAL	-0.245 (p<0.001)	-0.273 (p<0.0001)
Ítems de ESTRATEGIA SUPERFICIAL	-0.119 (p<0.05)	-0.145 (p<0.05)
Ítems de AUTORREGULACIÓN	0.195 (p<0.01)	0.233 (p<0.001)

Nota. *r* de Pearson. Se destacan en negrita las correlaciones superiores a 0.20.

Para completar la imagen anterior, analizando una por una las preguntas que valoran enfoques de aprendizaje, vamos a exponer los dos ítems de enfoques de aprendizaje con los que hay una correlación más importante, los dos únicos que tienen una correlación, positiva o negativa, superior a 0.30 con alguno de los tipos de examen. En ambos casos, la correlación alta es con el TBC, siendo mucho menor con los otros dos tipos de examen. Son el 29 (1° en el orden de la escala de Biggs) y el 31 (3° en la escala de Biggs):

El ítem 29 (1° en el orden de preguntas de la escala de Biggs) es de MOTIVO PROFUNDO: *Me doy cuenta de que estudiar me proporciona con frecuencia un sentimiento de profunda satisfacción personal.* Exponemos en la tabla 4.60 las correlaciones entre este ítem y la opinión del estudiante sobre la *eficacia docente* de cada tipo de examen.

TABLA 4.60

Correlaciones del ítem 29 (*motivo profundo*) con la opinión favorable a cada tipo de examen

Tabla de correlaciones del ítem 29 (MP)	r de Pearson
<i>Eficacia docente</i> del TBC	0.371 ($p < 0.0001$)
<i>Eficacia docente</i> del <i>test por aciertos</i>	0.12 ($p < 0.05$)
<i>Eficacia docente</i> de las <i>preguntas abiertas</i>	0.229 ($p < 0.001$)

Nota. Resaltamos en negrita el valor más elevado de los tres.

Hay una buena correlación con la opinión favorable al TBC y una correlación menor, moderada, con la preferencia con el examen de preguntas abiertas.

El ítem 31 (3° en el orden de preguntas de la escala de Biggs) es de MOTIVO SUPERFICIAL): *Mi objetivo es aprobar el curso haciendo el mínimo trabajo posible.*

TABLA 4.61.

Correlaciones del ítem 31 (superficial) con la *eficacia docente* percibida por el alumno en cada tipo de examen

Tabla de correlaciones del ítem 31 (MS)	r de Pearson
<i>Eficacia docente</i> del TBC	-0.330 ($p < 0.0001$)
<i>Eficacia docente</i> del test por aciertos	-0.10 NS
<i>Eficacia docente</i> de las preguntas abiertas	-0.08 NS

Nota. Resaltamos en negrita el valor más elevado de los tres.

Se aprecia que hay una considerable correlación negativa entre esta pregunta de motivo superficial y el tener una opinión favorable sobre la eficacia del TBC, sin que ocurra lo mismo con los otros dos tipos de examen.

Podemos, por consiguiente, afirmar que la opinión de que el TBC es un eficaz sistema de examen, que obliga a trabajar más al alumno y que le hace aprender más (ítems 6, 9, 16, 19 y 22) se da en mayor medida en los alumnos que tienen enfoques de aprendizaje profundo. Igualmente, los alumnos que no tienen esa opinión del TBC son, en mayor medida, aquellos que tienen enfoques de aprendizaje superficiales. Con los otros dos sistemas de examen no ocurre esto -aunque hay una muy ligera correlación entre *motivo superficial* y el tener la opinión positiva sobre el examen de preguntas abiertas-.

Hay una asociación entre enfoques profundos con una buena opinión sobre el TBC, y enfoques superficiales con peor opinión sobre el TBC. Los alumnos con mayor puntuación en enfoques profundos de aprendizaje son los que tienen una mejor opinión del TBC como sistema de examen (alta *eficacia docente*); los que tienen enfoques más superficiales de aprendizaje son los que tienen una peor opinión de este sistema de examen (baja *eficacia docente*).

- **Opinión sobre la *eficacia docente* de los distintos tipos de examen e ítems de autorregulación (49, 50, 51, 52 y 53).**

En cuanto a los 5 ítems que hemos incluido como medida de la autorregulación, sucede algo parecido a lo que pasaba con los de enfoque profundo de aprendizaje, aunque un poco menos marcado. La correlación de los ítems de *autorregulación* con la *opinión favorable al TBC* es moderada (0.23) mientras que es muy ligera con los otros dos tipos de examen (0.12 con el examen de tipo *test por aciertos* y 0.13 con el *de preguntas abiertas*). La autorregulación, en el aspecto estudiado, se correlaciona algo más con la opinión favorable al TBC que con la opinión favorable a los otros dos tipos de examen encuestados.

TABLA 4.62.

Correlación del agrupamiento de opinión favorable sobre la eficacia de cada tipo de examen y el agrupamiento de autorregulación

	TBC 6 +9 +16 +19 + 22	<i>Test por aciertos</i> 5+8+15+17+21	<i>Preguntas a desarrollar</i> 7+10+17+20+23
Ítems de AUTORREGULACIÓN	0.233 (p<0.001)	0.12 (p<0.05)	0.13 (p<0.05)

Nota. Se resalta en negrita la correlación mayor; r de Pearson.

- **Correlaciones de las opiniones sobre *eficacia docente* de cada examen con otros ítems de la encuesta**

Exponemos en la tabla 4.63 la correlación de la *eficacia docente* de cada examen con otras preguntas de la encuesta (solo ponemos las correlaciones significativas, que son la superiores a 0.1186; los datos completos están en el anexo G):

TABLA 4.63.

Otras correlaciones de los agrupamientos de «eficacia docente» de cada examen.

	TBC 6 + 9+16+ 19 + 22	Test por aciertos 5+8+15+ 17+21	Preguntas a desarrollar 7+10+17+ 20+23
Ítem 2 - Examen que <i>más te gusta</i> : test por aciertos	NS	0.276 (p<0.001)	0.13 (p<0.05)
Ítem 3 - Examen que <i>más te gusta</i> : TBC	0.255 (p<0.001)	-0.1577 (p<0.01)	NS
Ítem 4 - Examen que <i>más te gusta</i> : preguntas abiertas	NS	NS	0.224 (p<0.001)
Ítem 11 - Examen que obliga a <i>memorizar</i> : test por aciertos	0.13 (p<0.05)	0.258 (p<0.001)	0.265 (p<0.001)
Ítem 12 - Examen que obliga a <i>memorizar</i> : El TBC	0.337 (p<0.0001)	NS	0.354 (p<0.0001)
Ítem 13 - Examen que obliga a <i>memorizar</i> : preguntas abiertas	0.218 (p<0.001)	0.280 (p<0.001)	NS
Ítem 24 - En el examen que hice (TBC) <i>escogí bien mi grado de seguridad</i>	0.170 (p<0.01)	NS	NS
Ítem 26 - <i>Me estresa mucho hacer un examen de tipo test por aciertos</i>	NS	0.263 (p<0.001)	NS
Ítem 27 - <i>Me estresa mucho hacer un examen de tipo TBC</i>	0.159 (p<0.01)	0.173 (p<0.01)	NS
Ítems 15+ 18 + 21 - Validez autopercebida del test por aciertos	0.310 (p<0.001)	≈ 1	NS
Ítems 16+ 19 + 22 - Validez autopercebida del TBC.	≈ 1	0.313 (p<0.001)	NS
Ítems 17+ 20 + 23 - Validez autopercebida del examen de preguntas abiertas.	NS	NS	≈ 1

Nota. Solo ponemos las cifras de las correlaciones cuando son significativas (p> 0.119)

En la tabla 4.63 se observa que hay una correlación moderada entre la predilección por un examen y la opinión de que ese tipo de examen es un buen sistema de evaluación (r oscila entre 0.224 y 0.276), hecho que era previsible. Los alumnos que prefieren el TBC tienden a tener una opinión algo peor del test por aciertos (r de -0.158).

La correlación entre opinar que un examen obliga al alumno a «memorizar más» y la buena opinión sobre las cualidades de un examen, tal como se ve en este cuadro, es compleja: cuando el sujeto da una alta puntuación en «validez» o «eficacia docente» a alguno de los dos sistemas de tipo test (el TBC o el test *por aciertos*) tiende a opinar

que tanto ese tipo de test que considera bueno como el examen de *preguntas abiertas* le obligan a memorizar mucho (correlaciones de 0.337 y 0.218 en el TBC, de 0.258 y 0.280 en el *test por aciertos*); por el contrario, cuando el alumno manifiesta que el examen de *preguntas abiertas* es un buen sistema de evaluación, tiende a pensar que son los exámenes de tipo test los que le obligan a memorizar mucho (r de 0.354 para el TBC y de 0.265 para el test por aciertos). La consecuencia que se puede sacar de estas correlaciones es la de que el alumno que tiene una buena opinión de un sistema de test (sea TBC o *por aciertos*) tiende también a pensar que es un sistema de examen que obliga a memorizar mucho; por el contrario, cuando los alumnos dan una alta puntuación a la validez y eficacia del examen de *preguntas abiertas*, no lo asocian con el hecho de que ese examen obligue a memorizar mucho. Entre aquellos alumnos que muestran su predilección por los test, la opinión de que obligue a memorizar mucho no es sentida como una cualidad negativa, sino que ambos elementos, obligar a memorizar y ser un buen sistema de evaluación, se tienden a ver como relacionados. Eso no ocurre en los alumnos que otorgan una alta puntuación a los exámenes de *preguntas abiertas*; ahí no hay correlación entre «ser un buen sistema de evaluación» y «obligar a memorizar más».

4.2.1.6. RELACIÓN ENTRE EL SISTEMA DE EXAMEN Y LA GENERACIÓN DE ESTRÉS AL ALUMNO

Para valorar este aspecto se incluye una única pregunta, en los ítems 26 (test por aciertos), 27 (TBC) y 28 (preguntas abiertas):

Pregunta (ítems 26, 27 y 28): *¿Te sientes muy estresado antes de hacer un examen?*

Encuesta de 2013 y 2014 (272 alumnos)

Los estadísticos básicos de estos tres ítems -26, 27 y 28- muestran en la tabla 4.64. En la tabla 4.65 se exponen las significaciones estadísticas de las diferencias de las medias.

TABLA 4.64.

Puntuaciones medias de los estudiantes al estrés que perciben con cada tipo de examen

	Test por aciertos	TBC	Preguntas abiertas
MEDIA (de 1 a 6)	4.31	4.81	4.92
Desviación estándar	1.31	1.18	1.29
Error típico de la media	0.09	0.08	0.08
Media sobre 10	6.62	7.62	7.84

Nota. Resultados detallados en el anexo F.

TABLA 4.65.

Significación estadística de las diferencias (t de Student)

	Test por aciertos	TBC
TBC	p<0.001	–
Preguntas abiertas	p<0.001	p<0.05
F = 56.167	P<0.001	Eta ² = 0.172

En general, los alumnos suelen reconocer un nivel ligeramente mayor de estrés con los tipos de examen que les obligan a estudiar más, a entender mejor lo que estudian y a memorizar más, es decir, manifiestan un nivel de estrés mayor con el examen de preguntas abiertas y con el TBC que con el test por aciertos.

Los participantes opinan que el TBC les genera más estrés que el test *por aciertos* pero prácticamente lo mismo, solo ligeramente menos, que el examen de *preguntas abiertas*.

La correlación de estos tres ítems entre sí se expresa en la tabla 4.66:

TABLA 4.66.

Tabla de correlaciones de los ítems 26, 27 y 28, de estrés, con los exámenes

	Estrés con TBC	Estrés con <i>test por aciertos</i>	Estrés <i>preguntas abiertas</i>
Estrés con TBC	1		
Estrés <i>test por aciertos</i>	0.810	1	
Estrés <i>preguntas abiertas</i>	0.704	0.609	1

Nota. *r* de Pearson. Todas las correlaciones son significativas, $p < 0.0001$

Las altas correlaciones indican que cuando un alumno reconoce un alto nivel de estrés con un tipo de examen suele reconocer también alto nivel de estrés con los otros dos tipos. Igualmente, los alumnos que se reconocen poco estresados, lo suelen reconocer para los tres exámenes.

- **Correlación del estrés ante los exámenes y el género.**

Buscando las correlaciones de esta pregunta sobre el nivel de estrés que producen los exámenes, encontramos una correlación de moderada/grande magnitud con el género de los alumnos. Esto lo exponemos en la tabla 4.67.

TABLA 4.67.

Correlación estrés – género.

	Correlación con género femenino	Significación
Estrés con el test por aciertos	$r = 0.341$	<0.0001
Estrés con el TBC	$r = 0.318$	<0.0001
Estrés con preguntas abiertas	$r = 0.238$	<0.001

Nota. *r* de Pearson; extraído del anexo G.

El estudio de correlaciones de los ítems 26, 27 y 28 nos ha indicado que existe correlación entre el género del estudiante y el estrés autopercebido, correlación positiva con el género femenino y negativa con el masculino. Para completar ese dato presentamos el cuadro de estadísticos básicos que compara las dos cohortes, mujeres y hombres, en cuanto a su nivel de estrés autopercebido (tabla 4.68).

TABLA 4.68

Estrés de cada género con los tres tipos de examen. Ítem 26: **test por aciertos**.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	215	4.64	1.296	0.088	0.90	5.853	<0.01	5.275	<0.001
varón	55	3.42	1.583	0.214					

Nota. Enunciado del ítem: *Me siento muy estresado ante un examen de test convencional, sin valorar el grado de seguridad.*

TABLA 4.69.

Estrés de cada género con los tres tipos de examen. Ítem 27: **TBC**.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	215	5.08	1.112	0.076	0.83	18.96	<0.001	4.381	<0.001
varón	55	4.05	1.649	0.222					

Nota. Enunciado del ítem: *Me siento muy estresado ante un examen de test teniendo que indicar mi grado de seguridad.*

TABLA 4.70.

Estrés de cada género con los tres tipos de examen. Ítem 28: **preguntas abiertas**.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	215	5,20	1.173	0.080	0.60	14.28	<0.001	3.392	<0.001
varón	55	4.44	1.561	0.210					

Nota. Enunciado del ítem: *Me siento muy estresado ante un examen de tipo desarrollar un tema.*

En cuanto al nivel de estrés que los estudiantes reconocen que les generan los exámenes, las estudiantes de género femenino reconocen mayor estrés con los exámenes que los de género masculino. Esto sucede con los tres tipos de examen, pero en mayor medida con los exámenes de tipo test. La diferencia en el nivel de estrés entre ambos géneros es notable y significativa (tamaño del efecto entre un 60 % y un 90 % de la desviación típica, $p < 0.001$).

- **Correlación del estrés generado por los exámenes con los demás ítems**

Otras correlaciones de la edad, menos elevadas pero significativas, se expresan en la siguiente tabla (solo incluimos aquellas preguntas que tienen alguna correlación significativa mayor de 0,12, con $p < 0.05$; tabla 4.71):

TABLA 4.71.

Correlaciones de los ítems que miden estrés con cada examen y los otros ítems de la encuesta (anexo G)

	Estrés con el TBC (ítem 27)	Estrés con el test por aciertos (Ítem 26)	Estrés con preguntas abiertas (Ítem 28)
EDAD	-0.06 NS	-0.05 NS	-0.12 (p<0.05)
Nota obtenida en el examen (TBC)	-0.172 (p<0.01)	-0.163 (p<0.01)	-0.11 NS
Ítem 3 - Examen que <i>más te gusta</i> : TBC abiertas	NS	-0.165 (p<0.01)	NS
Ítem 4 - Examen que <i>más te gusta</i> : preguntas abiertas	NS	NS	-0.15 (p<0.05)
Ítem 5 - Examen que obliga a estudiar más: test por aciertos	0.14 (p<0.05)	0.172 (p<0.01)	NS
Ítem 6 - Examen que obliga a estudiar más: El TBC	0.177 (p<0.01)	NS	NS
Ítem 7 - Examen que obliga a estudiar más: preguntas abiertas	NS	NS	0.184 (p<0.01))
Ítem 8 - Examen que obliga a entender mejor: test por aciertos	NS	0.15 (p<0.05)	NS
Ítem 9 - Examen que obliga a entender mejor: El TBC	0.159 (p<0.01)	NS	NS
Ítem 12 - Examen que obliga a memorizar más: El TBC	0.14 (p<0.05)	NS	NS
Ítem 13 - Examen que obliga a memorizar más: preguntas abiertas	0.14 (p<0.05)	0.15 (p<0.05)	0.14 (p<0.05)
Es un válido sistema de examen: El test por aciertos. Ítems 15+18 +21	0.15 (p<0.05)	0.258 (p<0.001)	NS
Es un eficaz sistema de examen: El test por aciertos. Ítems 5+8+15+18 +21	0.173 (p<0.01)	0.263 (p<0.001)	NS
Ítem 35 «Cuando veo que la asignatura no me resulta muy interesante, trabajo lo menos posible». (motivo superficial)	-0.221 (p<0.001)	-0.182 (p<0.01)	-0.180 (p<0.01)
Ítem 36 «Aprendo las cosas repitiéndolas hasta que me las sé de memoria, aunque no las entienda» (estrategia superficial)	0.218 (p<0.001)	0.259 (p<0.001)	0.175 (p<0.01)
ESTRATEGIA SUPERFICIAL (32, 36, 40, 44, 48)	0.13 (p<0.05)	0.12 (p<0.05)	0.159 (p<0.01)

Nota. Se exponen solo las correlaciones con los ítems que tienen alguna correlación significativa (p<0.05, al menos). r de Pearson.

Se encuentra una correlación moderada entre pensar que el test convencional *por aciertos* es un buen sistema de examen y manifestar un alto nivel de estrés con ese tipo de examen. Hay una correlación significativa, pero muy débil, entre las estrategias superficiales de estudio y el nivel de estrés declarado en los exámenes. Por último, los

alumnos que en su estrategia de estudio reconocen una estrategia de memorización sin comprensión (ítem 36) declaran tener mayor estrés ante todos los tipos de examen, especialmente los de tipo test (r de 0'26, 0'22 y 0'18 en exámenes de *test por aciertos*, TBC y *preguntas abiertas*).

4.2.2. RELACIÓN ENTRE EL SISTEMA DE EXAMEN Y LOS ENFOQUES DE APRENDIZAJE, PROFUNDOS Y SUPERFICIALES

En este apartado intentamos comprobar si el test basado en el grado de confianza identifica mejor los enfoques de aprendizaje, profundos o superficiales, que el test tradicional con puntuación basada en el número de aciertos.

Para poder responder a esta pregunta hemos incluido en la encuesta la escala sobre enfoques de aprendizaje de Biggs de 20 ítems (R-SPQ-2F). Este cuestionario analiza los enfoques de aprendizaje del alumno dividiéndolos en cuatro categorías: Motivación profunda, estrategia profunda, motivación superficial y estrategia superficial; cada categoría cuenta con cinco preguntas del cuestionario. Hemos utilizado la traducción al español validada por Juan Carlos Torre (2007).

- Escala de valoración de las estrategias de aprendizaje de Biggs de 20 ítems (R-SPQ-2F):

10 ítems de aprendizaje profundo (5 de *motivos profundos*, 5 de *estrategias profundas*)

10 ítems de aprendizaje superficial (5 de *motivos superficiales*, 5 de *estrategias superficiales*)

Consistencia interna de la escala: Nosotros hemos obtenido los siguientes valores al calcular la fiabilidad o consistencia interna del test de Biggs en nuestra encuesta:

- Consistencia interna de la escala de enfoque profundo: α de Cronbach de **0.804**
 - Subescala de motivo profundo: α de Cronbach de 0.657
 - Subescala de estrategia profunda: α de Cronbach de 0.717
- Consistencia interna de la escala de enfoque superficial: α de Cronbach de **0.760**
 - Subescala de motivo superficial: α de Cronbach de 0.580
 - Subescala de estrategia superficial: α de Cronbach de 0.668

Las puntuaciones medias que los alumnos dan a los enfoques de aprendizaje, superficial y profundo en la encuesta se exponen en la tabla 4.72 y 4.73.

TABLA 4.72.

Tabla de estadísticos básicos de los ítems que miden enfoques profundos.

	Total ítems de enfoques profundos	Ítems de motivación profunda (29,33,37,41, 45)	Ítems de estrategia profunda (30, 34, 38, 42, 46)
Puntuación media	38.63 [38.1-39.1]	20.51 [20.25-20.76]	18.12 [17.90-18.36]
Desviación típica	7.058	3.905	3.928
Porcentaje sobre puntuación máxima	57.2 %	62.0 %	52.6 %

Nota. Se expresa entre corchetes tras la media el intervalo de confianza de la media. N de 272 participantes.

TABLA 4.73.

Tabla de estadísticos básicos de los ítems que miden **enfoques superficiales**.

	Total ítems de enfoques superficiales.	Ítems de motivación superficial (31, 35, 39, 43, 47)	Ítems de estrategia superficial (32, 36, 40, 44, 48)
Puntuación media	25.50 [25.01-25.59]	11.50 [11.29-11.71]	14.00 [13.74-14.26]
Desviación típica	6.924	3.523	4.262
Porcentaje sobre puntuación máxima	31.0 %	26.0 %	36.0 %

Nota. Se expresa entre corchetes tras la media el intervalo de confianza de la media. N de 272 participantes.

Comparando los enfoques profundos con los superficiales obtenemos las siguientes tres tablas (4.74, 4.75 y 4.76):

Tabla 4.74

ENFOQUE	Media	δ	Error T.	d	F	p	t	p
profundo	38.63	7.058	0.487	1.87	299.97	<0.0001	19.1	<0.0001
superficial	25.51	6.924	0.471					

Nota. 272 participantes. 10 ítems en cada enfoque

Tabla 4.75

MOTIVO	Media	δ	Error T.	d	F	p	t	p
profundo	20.51	3.905	0.237	2.42	796.43	<0.0001	25.37	<0.0001
superficial	11.50	3.523	0.214					

Nota. 272 participantes. 5 ítems en cada enfoque

Tabla 4.76

ESTRATEGIA	Media	δ	Error T.	d	F	p	t	p
profundo	18.12	3.928	0.238	1.01	138.13	<0.0001	10.84	<0.0001
superficial	14.00	4.262	0.258					

Nota. 272 participantes. 5 ítems en cada enfoque

Los alumnos dan puntuaciones más altas a los ítems de *enfoque profundo* que a los de *enfoque superficial*. Las diferencias entre las puntuaciones de los enfoques profundos y superficiales son en todos los casos estadísticamente significativas ($p < 0.0001$).

Es llamativo que los de *motivo profundo* reciban más puntuación que los de *estrategia profunda*, lo contrario que ocurre en los *enfoques superficiales*; aquí los ítems de *motivación superficial* reciben menos puntuación que los de *estrategia superficial*. En la opinión del autor, al responder a las preguntas sobre *motivación*, que son preguntas sobre actitudes, siempre subjetivas, los alumnos tienen tendencia a expresar lo que *debería ser* más que lo que es; por el contrario, al responder a las preguntas sobre *estrategias*, que son preguntas sobre actividades concretas, los alumnos tienden a ser más objetivos, con menos tendencia al sesgo hacia lo que *debería ser*.

Consideramos que los estudiantes tienden a ser más sinceros cuando se les interroga sobre estrategias que sobre motivaciones. Esto tiene trascendencia al establecer correlaciones con otras preguntas, como luego veremos.

4.2.2.1. CORRELACIONES DE LOS ENFOQUES PROFUNDOS DE APRENDIZAJE

Vamos a exponer las correlaciones de los enfoques profundos de aprendizaje con los otros ítems de la encuesta (los datos completos figuran en el anexo G, 11 y 12).

1.-En la tabla 4.77 exponemos la correlación entre los ítems de **motivo profundo** con los resultados del examen, tanto corregidos mediante el sistema de TBC como mediante el test basado en número de aciertos.

TABLA 4.77.

Correlación motivo profundo–resultados en el examen.

MOTIVO PROFUNDO	Correlación con la nota en el TBC (r de Pearson)	Correlación con la nota por n.º de aciertos (r de Pearson)
Ítem 29	0.091	0.073
Ítem 33	-0.016	-0.024
Ítem 37	0.094	0.087
Ítem 41	0.292	0.246
Ítem 45	0.167	0.123
Suma de ítems (29+33+37+41+45)	0.181	0.145
Significación de la r	p<0.01	p<0.05

Nota. r de Pearson. La diferencia de las correlaciones de los dos de exámenes no es significativa (se ha utilizado la página SISA, Simple Interactive Analysis, <http://www.quantitativeskills.com/sisa/>, consultada el 11 de febrero de 2015)

2.-En la tabla 4.78 exponemos la correlación entre los ítems de **estrategia profunda** con los resultados del examen, tanto corregidos mediante el sistema de TBC como mediante el test basado en número de aciertos.

TABLA 4.78.

Correlación estrategia profunda–resultados en el examen.

ESTRATEGIA PROFUNDA	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Ítem 30	0.205	0.160
Ítem 34	0.172	0.130
Ítem 38	0.267	0.203
Ítem 42	0.104	0.068
Ítem 46	0.174	0.133
Suma de ítems (30+34+38+42+46)	0.268	0.202
Significación de la correlación	p<0.001	P<0.001

Nota. r de Pearson. La diferencia de las correlaciones de los dos de exámenes no es significativa (se ha utilizado la página SISA, Simple Interactive Analysis, <http://www.quantitativeskills.com/sisa/>, consultada el 11 de febrero de 2015)

Las puntuaciones de los alumnos se correlacionan con sus enfoques de aprendizaje de manera pequeña–moderada. Las correlaciones con los *motivos profundos* son pequeñas aunque significativas (r de 0.179 con p<0.01 con el TBC y 0.145 con p<0.05 con el test por aciertos). Las correlaciones con las *estrategias profundas* son algo mayores (con el TBC, la r es de 0.264 con p<0.001 y con la nota por aciertos la r es de 0.202, p<0.001).

Si establecemos la correlación con los 10 ítems de enfoques profundos de aprendizaje (tanto de *motivo* como de *estrategia*), obtenemos los datos que exponemos en la tabla 4.79.

TABLA 4.79.

Correlación entre los ítems de enfoque profundo y los resultados en los exámenes.

Correlación con la suma de todos los ítems de enfoques profundos(r de Pearson).	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Enfoque profundo (29 +30 +33 +34 +37 +38 +41 +42 +45 +46)	0.250	0.195
Significación de la correlación	p<0.001	p<0.01

Estas correlaciones nos parecen escasas para lo que esperábamos encontrar, pues cabría pensar que los alumnos con enfoques profundos de aprendizaje tuviesen una

tendencia más clara a tener notas más elevadas. Cuando desglosamos los resultados por años, analizando separadamente las encuestas pasadas el año 2013 y el año 2014 nos encontramos con una sorpresa: los resultados difieren notablemente de un año para otro.

Datos desglosados por años:

En la tabla 4.80 exponemos los datos correspondientes a las correlaciones de los **enfoques profundos** con los resultados académicos separados por años, de 2013 y 2014, junto con los datos conjuntos de ambos, expuestos en la tabla 4.73.

TABLA 4.80.

Resultados separados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **enfoque profundo** (motivo + estrategia) y los resultados en los exámenes

ITEMS 29, 30, 33, 34, 37, 38, 41, 42, 45 y 46 (10 ítems)		
Correlación con la suma de los ítems de enfoque profundo (r de Pearson)	Correlación con la nota en el TBC	Correlación con la nota por n° de aciertos
Grupo de 2013 (134 sujetos)	0.375 (p<0.001)	0.327 (p<0.001)
Grupo de 2014 (138 sujetos)	0.123 (NS)	0.069 (NS)
Grupo total (272 sujetos)	0.250 (p<0.001)	0.195 (p<0.01)

Nota. Las correlaciones se calculan según la r de Pearson.

En la tabla 4.81 exponemos las correlaciones resultados académicos/*motivos profundos* en los años 2013, 2014 y conjunto –estos últimos ya expuestos en la tabla 4.77–.

TABLA 4.81.

Resultados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **motivo profundo** y los resultados en los exámenes

Correlación con la suma de los ítems de motivo superficial	Correlación con la nota en el TBC	Correlación con la nota por n° de aciertos
Grupo de 2013 (134 sujetos)	0.264 (p<0.01)	0.229(p<0.01)
Grupo de 2014 (138 sujetos)	0.098(NS)	0.068 (NS)
Grupo total (272 sujetos)	0.181(p<0.01)	0.145(p<0.05)

Nota. Las correlaciones se calculan según la r de Pearson.

En la tabla 4.82 exponemos las correlaciones resultados académicos/*estrategias profundas* en los años 2013, 2014 y conjunto –estos últimos ya expuestos en la tabla 4.78.

TABLA 4.82
Resultados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **estrategia profunda** y los resultados en los exámenes

Correlación con la suma de los ítems de motivo superficial	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Grupo de 2013 (134 sujetos)	0.410 (p<0.0001)	0.364 (p<0.0001)
Grupo de 2014 (138 sujetos)	0.123 (NS)	0.054 (NS)
Grupo total (272 sujetos)	0.268 (p<0.001)	0.202 (p<0.001)

Nota. Las correlaciones se calculan con la r de Pearson.

Vemos que hay una notable diferencia entre los dos años, 2013 y 2014, en las correlaciones entre los resultados del examen y las respuestas a los ítems de enfoques profundos; son correlaciones moderadas y significativas en el grupo de 2013 y no hemos podido encontrar correlación en el de 2014 (debido al efecto de uno de los grupos de ese año, cuyas respuestas fueron en sentido opuesto a los demás grupos, rebajando la correlación). Con los resultados globales, 2013 y 2014, las correlaciones son menores que en los de 2013, pero son significativas por el mayor número de sujetos.

Los alumnos con enfoques profundos de aprendizaje obtienen mejores resultados académicos de media que los que manifiestan enfoques superficiales. En todos los casos la correlación es mayor con los resultados del TBC que con el simple número de aciertos. Este punto lo estudiaremos con más detalle cuando veamos el ítem «TBC–aciertos» más adelante.

4.2.2.2. CORRELACIONES ENTRE EL RENDIMIENTO ACADÉMICO Y LOS ENFOQUES DE APRENDIZAJE SUPERFICIALES

En la tabla 4.83 se exponen las correlaciones (r de Pearson) entre las puntuaciones a los ítems de enfoque superficial de aprendizaje -10 ítems en total- y los resultados en el examen, según TBC y según *test por aciertos*.

TABLA 4.83.

Resultados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **enfoque superficial** (motivo + estrategia) y los resultados en los exámenes

ITEMS 31, 32, 35, 36, 39, 40, 43, 44, 47 y 48 (10 ítems)		
Correlación con la suma de los ítems de enfoque superficial	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Grupo de 2013 (134 sujetos)	-0.230 (p<0.01)	-0.236 (p<0.01)
Grupo de 2014 (138 sujetos)	-0.215 (p<0.05)	-0.206 (p<0.05)
Grupo total (272 sujetos)	-0.208 (p<0.001)	-0.205 (p<0.001)

Nota. La correlación se calcula mediante la r de Pearson.

En la tabla 4.84 se exponen las correlaciones (r de Pearson) entre las puntuaciones a los ítems de **motivo superficial** de aprendizaje -5 ítems en total- y los resultados en el examen, según TBC y según *test por aciertos*.

TABLA 4.84.

Resultados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **motivo superficial** y los resultados en los exámenes

Correlación con la suma de los ítems de motivo superficial	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Grupo de 2013 (134 sujetos)	-0.190 (p<0.05)	-0.200 (p<0.05)
Grupo de 2014 (138 sujetos)	-0.136 (NS)	-0.145 (NS)
Grupo total (272 sujetos)	-0.162 (p<0.01)	-0.169 (p<0.01)

Nota. La correlación se calcula mediante la r de Pearson.

En la tabla 4.85 se exponen las correlaciones (r de Pearson) entre las puntuaciones a los ítems de **estrategia superficial** de aprendizaje -5 ítems en total- y los resultados en el examen, según TBC y según *test por aciertos*.

TABLA 4.85.

Resultados de 2013, 2014 y conjunto. Correlaciones entre los ítems de **estrategia superficial** y los resultados en los exámenes

Correlación con la suma de los ítems de estrategia superficial	Correlación con la nota en el TBC	Correlación con la nota por nº de aciertos
Grupo de 2013 (134 sujetos)	-0.216 (p<0.05)	-0.221 (p<0.05)
Grupo de 2014 (138 sujetos)	-0.236 (p<0.01)	-0.215 (p<0.05)
Grupo total (272 sujetos)	-0.228 (p<0.001)	-0.216 (p<0.001)

Nota. La correlación se calcula mediante la r de Pearson.

En el caso de los *enfoques superficiales*, hay una ligera o moderada correlación negativa entre la alta puntuación a los enfoques superficiales de aprendizaje y los peores resultados académicos (r de -0.23 aproximadamente). La correlación negativa es parecida tanto con motivos como con estrategias (r de -0.208 en el primer caso, y de -0.205 en el segundo), con índices muy parecidos tanto con la nota obtenida en el TBC como con la nota de test por aciertos. A diferencia de lo que vimos que sucedía con los enfoques profundos, no hay casi diferencia entre los dos grupos, el de 2013 y el de 2014. En los ítems de *motivación superficial* las correlaciones son pequeñas (-0.162 y -0.169 para los dos tipos de examen analizados) y con poca diferencia entre 2013 y 2014; en cuanto a los ítems de enfoque superficial no hay diferencias entre ambos años y en todos los casos, las correlaciones son moderadas (entre 0.216 y 0.236), así como significativas.

Las respuestas de los alumnos a las preguntas sobre enfoques de aprendizaje se correlacionan con los resultados académicos; los enfoques profundos se asocian a mejores resultados y los superficiales a peores. Esto sucede en mayor medida con las «estrategias» que con los «motivos», con una intensidad moderada, algo inferior a lo que esperábamos.

Que las correlaciones sean algo menores de lo esperado se debe a que los resultados de las encuestas del año 2014 no mostraban correlación entre los ítems de enfoques profundos y las notas en el examen.

4.2.2.3. DIFERENCIAS EN LAS CORRELACIONES «ENFOQUE DE APRENDIZAJE– RESULTADOS ACADÉMICOS» ENTRE EL TBC Y EL TEST *POR ACIERTOS*

Como se ve en las tablas 4.77, 4.78, 4.84 y 4.85, los buenos resultados del examen se correlacionan moderadamente con las *estrategias* profundas de aprendizaje. La correlación es más alta en el caso del test *basado en la confianza* (r de 0.268) que con el test *por aciertos* (r de 0.202), pero la diferencia de correlaciones no es significativa. La correlación es menor con los *motivos* profundos de aprendizaje (r de 0.181 y 0.147), correlación escasa pero significativa en ambos casos ($p < 0.01$ en el primer caso, $p < 0.05$ en el segundo).

Con los enfoques superficiales no se encuentran diferencias con los resultados obtenidos mediante ambos tipos de test.

Dada la trascendencia que, de acuerdo con los objetivos del presente trabajo, tiene esta diferencia de correlaciones con cada tipo de test, estudiamos este punto añadiendo a la encuesta un nuevo parámetro; a ese nuevo ítem lo llamamos «TBC–aciertos». Si aparece una correlación significativa de algún tipo de enfoque de aprendizaje con este parámetro –«TBC-aciertos»- eso querrá decir que el TBC premiará –o castigará, si la correlación es negativa- el enfoque de aprendizaje correspondiente.

4.2.2.4. ÍTEM «TBC–ACIERTOS»

Para ver directamente la correlación entre la habilidad en manejar el TBC y los enfoques de aprendizaje, y cuantificarla mejor, hemos añadido un ítem específico a la encuesta; este ítem es la diferencia que el alumno tiene entre la nota que obtuvo según el sistema de puntuación basado en la seguridad, TBC, y la que habría obtenido si se hubiera puntuado solo según el número de aciertos. A ese ítem le hemos llamado «*TBC–aciertos*» e indica **la mayor o menor habilidad que el alumno tiene con este nuevo sistema de examen**, Así, los alumnos que se desenvuelven bien con él tienen una puntuación «*TBC–aciertos*» más alta que los alumnos que manejan mal el sistema, que tienen menos nota con el TBC de la que hubieran obtenido si el examen se hubiera puntuado por aciertos.

Las notas medias obtenidas en cada tipo de examen de test, y la puntuación *TBC–aciertos* se expresan en la siguiente tabla:

TABLA 4.86.
Estadísticos de los resultados en el examen (TBC, *nota por aciertos*, TBC-*aciertos*).

	Nota con TBC	Nota por aciertos	TBC–aciertos
Nota media	6.16	6.22	0.04
Desviación típica	1.95	1.88	0.59
Mínimo – Máximo	0.4 -10.87	0 - 10	-1.75, +1.29

Nota. N = 270 sujetos. Datos completos en el anexo F.

En la tabla 4.86 se observa que la nota media de los 270 alumnos es muy parecida, casi igual, en el TBC y la *nota por aciertos*, de forma que el *TBC–aciertos* medio es muy próximo a 0 (0.04). Pero se ve que hay importantes diferencias individuales entre ambas notas, y por lo tanto en el *TBC–aciertos*, que tiene una desviación típica de 0.59 y puede oscilar en un rango de 3.04, de -1.75 a 1.29, entre el caso mínimo y el máximo.

4.2.2.5. CORRELACIONES DEL ÍTEM «TBC-ACIERTOS»

Vamos a ver las principales correlaciones de este ítem «*TBC-aciertos*». Nos interesan especialmente sus correlaciones con los enfoques de aprendizaje.

De entrada no se correlaciona con el número de aciertos que el alumno tiene en el examen. La tabla 4.87 expone la correlación del TBC-aciertos con el número de aciertos en el examen (o con la *puntuación por acierto*, que es lo mismo)

TABLA 4.87.

Correlación entre la habilidad en el manejo del TBC y las notas por aciertos.

CORRELACIONES	<i>TBC-aciertos</i>
Nota por número de aciertos	-0.076 (NS)

Nota. La correlación se mide mediante la *r* de Pearson.

Es una correlación mínima y no significativa entre *TBC-aciertos* y el número de aciertos. Este dato significa que **la habilidad en el uso del TBC**, para obtener la máxima puntuación en él, **no se correlaciona con los conocimientos del alumno**, de los que depende el número de aciertos en el test. Esto es relevante, pues luego veremos que varias correlaciones importantes de este ítem lo son con otros que sí se correlacionan con los resultados en el examen; así podemos estar seguros de que cualquier correlación de un ítem con el «TBC-aciertos» es *independiente* de la que pueda tener con la nota en el examen.

El parámetro *TBC-aciertos* sí se correlaciona con la puntuación del examen según el grado de seguridad (*r* de Pearson de 0.220), como es de esperar, pues la habilidad en la utilización del TBC (medida por el ítem *TBC-aciertos*) es un componente de la nota. La nota TBC y la habilidad en la utilización del TBC –o acierto al escoger el nivel de confianza– son dos variables dependientes; por ese motivo no utilizamos, como parámetro de los resultados en el examen, la nota del TBC sino la *nota por aciertos*.

Otras correlaciones del ítem *TBC–aciertos*:

Las correlaciones de este ítem con las preguntas sobre la opinión del alumno sobre los diferentes tipos de examen han sido pequeñas en la mayoría de los casos, con excepción de la correspondiente al ítem 24 y a los **enfoques de aprendizaje**. A continuación las exponemos en la tabla 4.88; en ella ponemos solo las correlaciones significativas de $p < 0.05$ o mayor; los resultados completos se encuentran en el anexo G.

TABLA 4.88.
Correlaciones del *TBC–aciertos* con los otros ítems.

	<i>TBC–nota por aciertos</i>
Nota en el TBC	0.220 ($p < 0.001$)
Ítem 3: ¿Te gusta más el TBC?	0.168 ($p < 0.01$)
Ítem 6: El TBC me obliga a dedicar más tiempo al estudio	0.129 ($p < 0.05$)
Ítem 12: El TBC me obliga a memorizar más datos	0.133 ($p < 0.05$)
Ítem 22: El TBC me hace aprender más	0.124 ($p < 0.05$)
Ítem 23: El examen de preguntas abiertas me hace aprender más.	0.140 ($p < 0.05$)
Ítem 24: En el TBC suelo escoger bien mi grado de seguridad	0.252 ($p < 0.001$)
Ítem 25: Creo que este curso sacaré varios sobresalientes	0.173 ($p < 0.01$)
ENFOQUE PROFUNDO (29, 30, 33, 34, 37, 38, 41, 42, 45, 46)	0.196 ($p < 0.01$)
MOTIVO PROFUNDO (29, 33, 27, 41, 45)	0.129 ($p < 0.05$)
ESTRATEGIA PROFUNDA (30, 34, 38, 42, 46)	0.229 ($p < 0.001$)
AUTORREGULACIÓN : ítems 49, 50, 51, 52 y 53,	0.148 ($p < 0.05$)

Nota. La correlación se calcula mediante la *r* de Pearson. Entre paréntesis la significación estadística. Destacamos en negrita las correlaciones con $p < 0.01$. $N = 271$.

Como se ve en la tabla, las únicas correlaciones de mediana intensidad, superiores a 0.20, se dan con la nota en el TBC -que es una variable dependiente de *TBC–aciertos*-, con el ítem 24 y con la suma de ítems de **estrategia profunda**. En el ítem 24 al alumno se le pide que valore su acierto al escoger el grado de seguridad con el TBC; dado que el ítem «*TBC–aciertos*» mide de forma objetiva eso mismo, el acierto al escoger el grado de seguridad de la respuesta, es coherente que haya correlación entre ambos

ítems, el subjetivo (24) y el objetivo («*TBC–aciertos*»). A los alumnos se les había proporcionado previamente a la encuesta información detallada sobre los resultados de su examen, incluyendo la nota según el sistema TBC y según el test por aciertos, de manera que sabían exactamente en qué medida el TBC les había beneficiado o perjudicado.

Vamos a ver a continuación, separadamente y con el mayor detalle, los resultados de las correlaciones del ítem *TBC–aciertos* con los enfoques de aprendizaje.

4.2.2.6. CORRELACIONES ENTRE LA HABILIDAD EN EL MANEJO DEL TBC Y LOS ENFOQUES DE APRENDIZAJE DEL ALUMNO

A continuación veremos cuál es la correlación del *TBC–aciertos* con los resultados de la escala de Biggs de 20 ítems para valoración de los enfoques de aprendizaje. Primero veremos –en la tabla 4.89- la correlación entre el *TBC–aciertos* y los 10 ítems de *enfoque profundo*, sumando los de *motivo profundo* con los de *estrategia profunda*.

TABLA 4.89.

Correlaciones entre *TBC–aciertos* y factor «enfoque profundo»

Ítems de enfoque profundo (motivo y estrategia) 29+30+33+34+37+38+41+42+45+46	<i>TBC–aciertos</i>
Encuesta de 2013, 133 sujetos	0.232 (p<0.01)
Encuesta de 2014, 139 sujetos	0.172 (p<0.05)
Encuestas de 2013 y 2014, 272 sujetos	0.196 (p<0.01)

Nota. Las correlaciones se calculan mediante la r de Pearson.

Hay una correlación de ligera magnitud en el total de los dos años, cercana a 0.20 y significativa de p<0.01.

- **Correlaciones TBC-aciertos y *estrategia profunda*.**

A continuación, en la tabla 4.90, veremos las correlaciones con el grupo de 5 ítems que agrupamos en *estrategia profunda*.

TABLA 4.90.

Correlaciones entre *TBC-aciertos* y **estrategia profunda**.

Ítems de estrategia profunda: 30 + 34 + 38 + 42 + 46	TBC – aciertos
Encuesta de 2013, 133 sujetos	0.250 (p<0.01)
Encuesta de 2014, 139 sujetos	0.213 (p<0.05)
Encuestas de 2013 y 2014, 272 sujetos	0.229 (p<0.001)

Nota. Las correlaciones se calculan mediante la r de Pearson.

Vemos que hay una correlación de moderada magnitud en el total de los alumnos entre la habilidad para usar el TBC y las estrategias de aprendizaje profundo. En el conjunto de los 272 sujetos la r de esta correlación es de 0.229, con una alta significación estadística (p<0.001). En consecuencia, **el TBC identifica mejor las estrategias profundas de aprendizaje que el test convencional puntuado solo por el número de aciertos**. Con el término «identifica mejor» queremos decir que las premia con una puntuación más alta.

- **Correlaciones TBC-aciertos y *motivo profundo*.**

A continuación, en la tabla 4.91, veremos las correlaciones con el grupo de 5 ítems que agrupamos en *motivo profundo*:

TABLA 4.91.

Correlaciones entre *TBC-aciertos* y **motivo profundo**.

Ítems de motivo profundo: 29 + 33 + 37 + 41 + 45	TBC– aciertos
Encuesta de 2013, 133 sujetos	0.167 (p<0.05)
Encuesta de 2014, 139 sujetos	0.093 (NS)
Encuestas de 2013 y 2014, 272 sujetos	0.128 (p<0.05)

Nota. Las correlaciones se calculan mediante la r de Pearson.

Hay correlaciones muy débiles en 2013 y 2014, no significativas en el último año, y significativas en el conjunto de los dos años ($p < 0.05$). Las correlaciones de la *habilidad en desenvolverse con el TBC* y los *motivos profundos* son débiles, aunque significativas.

El TBC sirve para premiar las estrategias de aprendizaje profundas con mayor puntuación. Los alumnos que tienen mayor puntuación en estrategias profundas de aprendizaje son los que se manejan mejor con el TBC, obteniendo una nota más alta con él que con el *test por aciertos*.

4.2.2.7. CORRELACIONES ENTRE LOS ENFOQUES SUPERFICIALES DE APRENDIZAJE Y EL ÍNDICE *TBC-ACIERTOS*

Las correlaciones de la habilidad en el manejo del TBC y los enfoques de aprendizaje superficiales (motivo, estrategia y enfoque conjunto) las mostramos en la tabla 4.92, donde podemos observar que son mínimas y no significativas.

TABLA 4.92.

Correlaciones entre *TBC-aciertos* y enfoques superficiales

	Motivo superficial	Estrategia superficial	Enfoque superficial
2013, 133 sujetos	-0.015	-0.045	-0.036
2014, 139 sujetos	0.026	-0.069	-0.028
2013 y 2014, 272 sujetos	0.008	-0.059	-0.027

Nota. Ninguna de las correlaciones es significativa. se utiliza la *r* de Pearson.

Como se ve en la tabla, **no encontramos correlación entre la habilidad en el manejo del TBC y los enfoques de aprendizaje superficiales**, ni conjuntamente ni separándolos en enfoques profundos y superficiales.

4.2.3. ESTUDIO DE LOS RESULTADOS OBTENIDOS EN LOS ÍTEMS QUE VALORAN LA AUTORREGULACIÓN DEL ALUMNO

Hemos incluido en la encuesta un conjunto de 5 ítems que valoran la autorregulación, extraídos del cuestionario de autorregulación de Torres (2007). A continuación ponemos el enunciado de cada uno de estos ítems y, entre paréntesis, la subescala a la que pertenecen dentro de la escala global de autorregulación

Ítem 49.- Mientras estudio me doy cuenta de si voy cumpliendo o no mis objetivos (Control y verificación).

Ítem 50.- No utilizo siempre los mismos modos para trabajar/aprender los temas, sé cambiar de estrategia (Conciencia metacognitiva activa).

Ítem 51.- Cuando estoy estudiando una asignatura, sé exactamente qué quiero conseguir o aprender (Conciencia metacognitiva activa).

Ítem 52.- Cuando estoy estudiando, ante las dificultades, hago un esfuerzo mayor o cambio mi modo de estudiar (Control y verificación).

Ítem 53.- Cuando estoy trabajando una asignatura, intento identificar las ideas y conceptos que no comprendo (Control y verificación).

Vamos a mostrar, en primer lugar, el cuadro de correlaciones internas de este agrupamiento en la tabla 4.93.

TABLA 4.93.

Correlaciones entre los ítems de autorregulación.

	Ítem 49	Ítem 50	Ítem 51	Ítem 52	Ítem 53
Ítem 49	1				
Ítem 50	0.218 p<0.001	1			
Ítem 51	0.548 p<0.0001	0.345 p<0.0001	1		
Ítem 52	0.324 p<0.0001	0.394 p<0.0001	0.376 p<0.0001	1	
Ítem 53	0.437 p<0.0001	0.258 p<0.001	0.421 p<0.0001	0.496 p<0.0001	1

Nota. Correlaciones según r de Pearson. Datos completos en el anexo G. N = 272 sujetos.

La consistencia interna de este agrupamiento de 5 ítems es: **α de Cronbach de 0.735** (el cálculo se realiza mediante el SPSS, anexo H). Por lo tanto, el agrupamiento formado por los 5 ítems que hemos escogido para valorar la autorregulación tiene una **consistencia interna aceptable**. Lo consideramos un instrumento útil para valorar la autorregulación con una escala de solo 5 elementos.

En la tabla 4.94 exponemos las correlaciones internas de cada ítem con el agrupamiento.

TABLA 4.94.

Correlaciones de cada ítem de autorregulación con la suma del resto de ítems.

	Suma del resto de ítems del agrupamiento	Alfa de Cronbach si se excluye el elemento
Ítem 49	0.503	0.689
Ítem 50	0.404	0.751
Ítem 51	0.578	0.657
Ítem 52	0.545	0.673
Ítem 53	0.537	0.682

Nota. α total de 0.735. Datos obtenidos del estudio de componentes mediante el SPSS (anexo H).

Se puede apreciar en ambas tablas que el ítem 50 es el que menos se correlaciona con los demás, de forma que la fiabilidad del agrupamiento aumenta al excluirlo. El enunciado de este ítem era: «*No utilizo siempre los mismos modos para trabajar/aprender los temas, sé cambiar de estrategia* ». El grupo sería más homogéneo sin ese ítem, pero hemos decidido mantenerlo en el agrupamiento.

4.2.3.1. CORRELACIONES DE LOS ÍTEMS DE AUTORREGULACIÓN CON LOS RESULTADOS ACADÉMICOS

En primer lugar se muestra en la tabla 4.95 la correlación entre los 5 ítems de autorregulación y los resultados, tanto con el TBC como con el *test por aciertos* y con el índice *TBC–aciertos*.

TABLA 4.95.
Correlación entre la autorregulación y las puntuaciones en el examen.

ÍTEMS	AUTORREGULACIÓN (49+50+51+52+53)
Puntuación en el TBC	0.344 ($p < 0.0001$)
Puntuación en <i>test por aciertos</i>	0.310 ($p < 0.0001$)
<i>TBC–aciertos</i>	0.148 ($p < 0.05$)

Nota. La correlación se calcula mediante la r de Pearson.

Hay correlación moderada de los ítems de autorregulación escogidos con los resultados académicos, y ligera (r de 0.15) con la habilidad en la utilización del TBC (*TBC–aciertos*).

Analizando las correlaciones de los resultados académicos con cada uno de los ítems del agrupamiento «autorregulación», obtenemos la tabla 4.96.

TABLA 4.96.

Correlaciones de las puntuaciones del examen con cada ítem de autorregulación.

	Nota TBC	Aciertos	TBC–aciertos
Ítem 49	0.334 (p <0.0001)	0.297 (p <0.0001)	0.096 <i>ns</i>
Ítem 50	0.117 <i>ns</i>	0.105 <i>ns</i>	0.089 <i>ns</i>
Ítem 51	0.286 (p <0.001)	0.282 (p <0.001)	0.037 <i>ns</i>
Ítem 52	0.238 (p <0.001)	0.201 (p <0.01)	0.158 (p <0.01)
Ítem 53	0.303 (p <0.0001)	0.264 (p <0.001)	0.161 (p <0.01)
Suma de los 5 ítems (49+50+51+52+53)	0.344 (p <0.0001)	0.310 (p <0.0001)	0.148 (p <0.05)

Nota. *r* de Pearson. Para 272 sujetos los niveles de significación de *r* son: 0.119, $p < 0.05$; 0.157, $p < 0.01$; 0.202, $p < 0.001$.

La escala de autorregulación se correlaciona con el buen resultado en los exámenes con un nivel de *r* de Pearson intermedio, muy significativo - $p < 0.0001$ -, tanto si los exámenes han sido corregidos según el grado de seguridad (*r* de 0.344) o según el número de aciertos (*r* de 0.310). La correlación es ligeramente mayor con los resultados obtenidos en el TBC que con la corrección según el número de aciertos; esto sucede en cada uno de los ítems, aunque en tres de ellos la diferencia no sea significativa. Aunque la diferencia en la correlación con ambos sistemas de exámenes es pequeña, es, sin embargo, significativa, con un nivel de significación de $p < 0.05$, como se ve en la columna «TBC–aciertos» (*r* de 0.148). Si solo consideramos el conjunto de los tres ítems que miden el factor «control y verificación», la correlación con el índice TBC–aciertos se eleva a $r = 0.179$ ($p < 0.01$).

Por consiguiente, la autorregulación se correlaciona moderadamente con el resultado académico, y ligera pero significativamente con el buen desempeño o habilidad con el test basado en el grado de seguridad (TBC–aciertos).

Si a este cuadro de correlaciones de la autorregulación con los resultados académicos le añadimos dos ítems que muestran dos aspectos de la autoeficacia docente, el 24 (*en el TBC suelo escoger bien mi nivel de seguridad*: autoeficacia en relación al TBC) y el 25 (*creo que este curso conseguiré varios sobresalientes*: autoeficacia docente en general), obtenemos el cuadro que exponemos en la tabla 4.97.

TABLA 4.97.

Otras correlaciones de la autorregulación.

ÍTEMS	Autorregulación (49+50+51+52+53)
Nota en el TBC	0.344 (p <0.0001)
Nota según nº de aciertos	0.310 (p <0.0001)
TBC–aciertos	0.148 (p<0.05)
Ítem 24 –En el TBC «suelo escoger bien el nivel de seguridad» (autoeficacia en TBC).	0.291 (p <0.0001)
Ítem 25 –«Creo que este curso tendré varios sobresalientes» (autoeficacia docente).	0.273 (p<0.001)

Nota. *r* de Pearson. Para 272 sujetos los niveles de significación de *r* son: 0.119, p<0.05; 0.157, p< 0.01; 0.202, p<0.001.

Vemos que también hay correlaciones positivas con todos estos ítems, de manera que la autorregulación se correlaciona bien tanto con los resultados objetivos del alumno en los exámenes como con los resultados esperados y la autoeficacia.

Incluyendo en un cuadro general todas las correlaciones de estos cinco factores - Notas en el examen TBC, autoeficacia, habilidad en el uso del TBC y autorregulación-, obtenemos la tabla 4.98.

TABLA 4.98.

Otras correlaciones significativas de los cinco factores incluyendo autorregulación.

	Nota TBC	TBC- <i>aciertos</i>	Ítem 24	Ítem 25	Autorregulación
Nota TBC	1				
TBC- <i>aciertos</i>	0.220 ($p < 0.001$)	1			
Ítem 24	0.385 ($p < 0.0001$)	0.250 ($p < 0.001$)	1		
Ítem 25	0.405 ($p < 0.0001$)	0.172 ($p < 0.01$)	0.344 ($p < 0.0001$)	1	
Autorregulación	0.344 ($p < 0.0001$)	0.148 ($p < 0.05$)	0.291 ($p < 0.0001$)	0.273 ($p < 0.0001$)	1

Nota. r de Pearson. Para 272 sujetos los niveles de significación de r son: 0.119, $p < 0.05$; 0.157, $p < 0.01$; 0.202, $p < 0.001$.

4.2.3.2. CORRELACIONES DE LOS ÍTEMS DE AUTORREGULACIÓN CON LOS ENFOQUES DE APRENDIZAJE

Las correlaciones entre la escala de autorregulación y los enfoques profundo y superficial de aprendizaje las mostramos en las siguientes tablas. En la tabla 4.99 para los enfoques profundo y superficial en general, y en la tabla 4.100 para los motivos y estrategias profundas y superficiales.

TABLA 4.99.

Correlación autorregulación – enfoques de aprendizaje.

	Enfoque profundo	Enfoque superficial
AUTORREGULACIÓN	0.471	-0.331
Significación	$p < 0.0001$	$p < 0.0001$

Nota. r de Pearson. Para 272 sujetos los niveles de significación de r son: 0.119, $p < 0.05$; 0.157, $p < 0.01$; 0.202, $p < 0.001$; 0.26, $p < 0.0001$.

TABLA 4.100.

Correlación autorregulación – motivos y estrategias de aprendizaje.

	Motivo profundo	Estrategia profunda	Motivo superficial	Estrategia superficial
AUTORREGULACIÓN	0.398	0.451	-0.308	-0.284
Significación	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$

Nota. r de Pearson. Para 272 sujetos los niveles de significación de r son: 0.119, $p < 0.05$; 0.157, $p < 0.01$; 0.202, $p < 0.001$; 0.26, $p < 0.0001$.

Como cabría esperar, los alumnos con mejor *autorregulación* tienden a tener enfoques de aprendizaje más profundos, tanto en *motivos* como en *estrategias*. Los que tiene peor autorregulación, en cambio, tienden claramente a tener enfoques de estudio superficiales.

4.2.3.3. CORRELACIONES DE OTROS ÍTEMS CON LA AUTORREGULACIÓN

Exponemos en la tabla 4.101 otras correlaciones significativas de la autorregulación (ordenadas de mayor a menor r), resumidas de los datos completos expuestos en el anexo G.

1º.- Correlaciones con ítems individualmente considerados:

TABLA 4.101.

Correlaciones de la autorregulación con ítems individuales.

Correlaciones mayores de 0.200	Correlación con autorregulación
Ítem 24 – En el test teniendo que indicar el grado de seguridad suelo escoger bien el nivel de seguridad de cada pregunta.	0.291 $p < 0.0001$
Ítem 22 – ¿El examen de tipo test basado en el grado de seguridad te hace aprender más?	0.288 $p < 0.0001$
Ítem 25 – Creo que este curso tendré varios sobresalientes.	0.273 $p < 0.0001$
Ítem 9 – ¿En qué medida te obliga a entender bien todo lo que estás estudiando el <i>test basado en el grado de seguridad</i> ?	0.259 $p < 0.001$
Ítem 8 – ¿En qué medida te obliga a entender bien todo lo que estás estudiando el examen tipo <i>test por aciertos</i> ?	0.257 $p < 0.001$
Otras correlaciones menores de 0.200	
Ítem 20 – ¿El examen de tipo <i>desarrollar un tema</i> te hace darte cuenta mejor de tus lagunas formativas?	0.165 $p < 0.01$
Ítem 7 – ¿En qué medida te obliga a dedicar más tiempo al estudio un examen de desarrollar un tema?	0.161 $p < 0.01$
Ítem 13 – ¿El examen de tipo desarrollar un tema te obliga a memorizar más?	0.156 $p < 0.05$
Ítem 3 – ¿Te gusta el test teniendo que indicar el grado de seguridad?	0.134 $p < 0.05$
Ítem 10 – ¿En qué medida te obliga a entender bien todo lo que estás estudiando el examen de desarrollar un tema?	0.134 $p < 0.05$

2°.- **Correlación de la autorregulación con los agrupamientos «validez» y «eficacia docente» de cada tipo de examen**

– En primer lugar, exponemos en la tabla 4.102 las correlaciones de la autorregulación con el agrupamiento de las tres preguntas que valoraban la *validez percibida por el alumno* de cada tipo de examen:

TABLA 4.102.

Correlaciones de la autorregulación con los agrupamientos que miden la «validez autopercebida» de cada tipo de examen.

	V.A. del TBC 16+19+22	V.A. del <i>test por aciertos</i> 15+17+21	V.A. de las <i>preguntas a desarrollar</i> 17+20+23
AUTORREGULACIÓN (49, 50, 51, 42, 43)	0.195 ($p < 0.01$)	0.037 NS	0.083 NS

Nota. Para 272 sujetos los niveles de significación de r son: 0.119 $\rightarrow p < 0.05$; 0.157 $\rightarrow p < 0.01$; 0.202 $\rightarrow p < 0.001$; 0.26 $\rightarrow p < 0.0001$ (r de Pearson). Señalamos en negrita la correlación significativa.

– En segundo lugar, en la tabla 4.103 exponemos la correlación de la autorregulación con el agrupamiento de las cinco preguntas que valoraban la *eficacia docente* de cada examen, según era percibida por el alumno.

TABLA 4.103.

Correlaciones de la autorregulación con los agrupamientos que miden la «eficacia docente» de cada tipo de examen y el esfuerzo que exige al alumno (r de Pearson).

	«Eficacia docente» del TBC 6 +9 +16 +19 + 22	«Eficacia docente» <i>Test por aciertos</i> 5+8+15+17+21	«Eficacia docente» <i>Preguntas a desarrollar</i> 7+10+17+20+23
AUTORREGULACIÓN (49, 50, 51, 42, 43)	0.233 ($p < 0.001$)	0.121 ($p < 0.05$)	0.129 ($p < 0.05$)

Nota. Para 272 sujetos los niveles de significación de r son: 0.119 $\rightarrow p < 0.05$; 0.157 $\rightarrow p < 0.01$; 0.202 $\rightarrow p < 0.001$; 0.26 $\rightarrow p < 0.0001$ (r de Pearson).

Como se ve, los alumnos con una autorregulación más elevada tienden a ser los que tienen mejor opinión del TBC como sistema eficaz de examen. Con los otros tipos de examen no hay correlación o esta es muy débil.

Estos son los ítems con los cuales la autorregulación tiene correlaciones significativas. Hemos destacado aquellos ítems en los que la correlación es superior a 0.201, lo que corresponde a una significación estadística de $p < 0.001$, de aquellas inferiores a esa cifra (entre 0.12 y 0.20), con correlación débil aunque significativa (0.1190 $p < 0.05$; 0.1567 $p < 0.01$). Estos ítems se pueden dividir en varios agrupamientos, lo que nos puede ofrecer una mejor información.

Así, podemos considerar que los ítems 3 (*Me gusta el test basado en el grado de seguridad*), 22 (*El examen de tipo test basado en el grado de seguridad me hace aprender más*) y 24 (*Me desenvuelvo bien con el test basado en el nivel de seguridad*) son ítems que indican una experiencia del alumno favorable con el TBC. Los tres tienen correlaciones positivas, de diferente magnitud, con la puntuación global obtenida. La correlación de la puntuación global en autorregulación con la suma de estos tres ítems de opinión positiva sobre el TBC se expone en la tabla 4.104.

TABLA 4.104.

Correlación autorregulación con ítems de «opinión positiva» sobre el TBC.

Ítems de <i>opinión positiva</i> sobre el TBC	Correlación con la puntuación en AUTORREGULACIÓN
3 (me gusta el TBC)	0.134 ($p < 0.05$)
22 (el TBC me hace aprender más)	0.288 ($p < 0.0001$)
24 (he estado acertado con el TBC)	0.291 ($p < 0.0001$)
SUMA DE 3 + 22 + 24	0.294 ($p < 0.0001$)

Nota. Para 272 sujetos los niveles de significación de r son: 0.119 $\rightarrow p < 0.05$; 0.157 $\rightarrow p < 0.01$; 0.202 $\rightarrow p < 0.001$; 0.26 $\rightarrow p < 0.0001$ (r de Pearson).

Los sujetos con mayor puntuación en autorregulación tienen tendencia a tener una opinión positiva del test basado en la confianza. Esto no sucede con ninguno de los otros dos tipos de examen

El siguiente grupo de ítems que tienen una correlación significativa con la *autorregulación* son los ítems 7, 8, 9, 10 y 13. Estos ítems se refieren a la cantidad de esfuerzo y tipo de estudio que cada tipo de examen supone. Ya los hemos visto en el estudio descriptivo de arriba, ahora vamos a ver la tabla 4.105, de correlaciones de estos ítems con la autorregulación.

TABLA 4.105.

Correlación de la autorregulación con ítems sobre el tipo de aprendizaje que induce cada uno de los tres tipos de examen.

Ítems de <i>opinión positiva</i> sobre los exámenes	Puntuación alta en <i>autorregulación</i>
5. <i>El test por aciertos te obliga a dedicar mucho tiempo al estudio</i>	NS
6. <i>El TBC te obliga a dedicar mucho tiempo al estudio</i>	NS
7. <i>El examen de preguntas abiertas te obliga a dedicar mucho tiempo al estudio</i>	0.161 (P<0.01)
8. <i>El test por aciertos te obliga a entender bien todo</i>	0.259 (P<0.001)
9. <i>El TBC te obliga a entender bien todo</i>	0.257 (P<0.001)
10. <i>El examen de preguntas abiertas te obliga a entender bien todo</i>	0.134 (P<0.05)
11. <i>El test por aciertos te obliga a memorizar más</i>	NS
12. <i>El TBC te obliga a memorizar más</i>	NS
13. <i>El examen de preguntas abiertas te obliga a memorizar más</i>	0.156 (P<0.05)

Nota. Para 272 sujetos los niveles de significación de r de Pearson: 0.119 → p<0.05; 0.157 → p<0.01; 0.202 → p<0.001; 0.26 → p<0.0001 (r de Pearson).

Este cuadro de correlaciones indica que los alumnos con elevada autorregulación tienden a ser más conscientes de que los exámenes de tipo test (en sus dos variedades) obligan a un mayor esfuerzo de comprensión. En medida algo menor, tienden a ser

conscientes de que el examen de tipo preguntas abiertas obliga a un mayor esfuerzo y estudio en general, tanto de comprensión como de memorización. En este último caso, las correlaciones, aunque significativas, son más débiles.

4.3. ESTUDIO DE LA INFLUENCIA DEL GÉNERO DE LOS ESTUDIANTES EN LOS RESULTADOS DE LA ENCUESTA

Hicimos un análisis de las diferencias en los resultados de la encuesta según el género de los alumnos. En total, hubo 271 alumnos que cumplimentaron el dato de su género: 216 mujeres y 55 varones. Para el estudio de correlaciones, transformamos el género en una variable numérica (1 para los varones, 0 para las mujeres*) y se analizó su correlación con todos los demás ítems. Hubo pocas correlaciones positivas. El cuadro de correlaciones lo exponemos, a continuación, en la tabla 4.106.

TABLA 4.106.

Correlaciones entre el género de alumno y los ítems de la encuesta.

	Correlación con varones	Correlación con mujeres	Significación de la r
Nota en el TBC	-0.024	0.024	NS
Diferencia TBC–aciertos	0.014	-0.014	NS
Ítem 26 - Estrés con el test por aciertos	-0.341	0.341	p<0.0001
Ítem 27 - Estrés con el TBC	-0.318	0.318	p<0.0001
Ítem 28 - Estrés con preguntas abiertas	-0.238	0.238	p<0.001
<i>Motivos profundos</i>	-0.099	0.010	NS
<i>Estrategias profundas</i>	-0.012	0.012	NS
<i>Motivos superficiales</i>	0.154	- 0.154	p<0.05
<i>Estrategias superficiales</i>	0.017	0.017	NS
<i>Autorregulación</i>	0.001	0.001	NS

Nota. Para 272 sujetos los niveles de significación de r son: 0.119 → p<0.05; 0.157 → p<0.01; 0.202 → p<0.001; 0.26 → p<0.0001 (r de Pearson). Se resaltan en negrita las correlaciones significativas.

La utilización de una tabla de correlaciones para ver si hay diferencias significativas entre mujeres y hombres en los distintos parámetros de la encuesta es un sistema eficiente y científicamente exacto de calcular el sesgo de género, sin tener que dividir el grupo total de sujetos en dos cohortes, varones/mujeres, y hacer un contraste de medias. En los ítems en los que no hay correlación significativa entre género y resultado no hacemos el estudio de cohortes, pues los valores que se obtuvieran no tendrían significación estadística. Hemos obtenido correlaciones significativas con el género en los ítems que valoran estrés ante el examen, como ya hemos analizado en el capítulo 4.2.1.5, y con la subescala de «motivo superficial», dentro de los enfoques de aprendizaje. No hubo correlación significativa del género con ningún otro de los ítems de la encuesta ni con los resultados académicos. En nuestra población estudiada, el género del alumno no se ha correlacionado con los resultados académicos.

Como sí hemos encontrado una ligera correlación entre enfoques de aprendizaje y género («motivo superficial»), pondremos en la tabla siguiente el cuadro de estadísticos básicos de los estudiantes de cada género.

TABLA 4.107.

Motivos profundos (suma de ítems 29, 33, 37, 41 y 45) en cada género. Estadísticos.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	216	20.68	3.92	0.267	0.24	0.226	>0.05	1.625	>0.05
varón	55	19.73	3.82	0.515			NS		NS

Nota. Suma de los valores de los ítems de motivo profundo.

TABLA 4.108.

Estrategia profunda (ítems 30, 34, 38, 42 y 46) en cada género. Estadísticos.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	216	18.13	3.98	0.271	0.03	0.097	>0.05	0.196	>0.05
varón	22	18.02	3.83	0.510			NS		NS

Nota. Suma de los valores de los ítems de estrategia profunda.

TABLA 4.109.

Motivo superficial (ítems 31, 35, 39, 43 y 47) en cada género. Estadísticos.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	216	11.22	3.572	0.244	-0.38	0.551	< 0.05	-2.544	< 0.05
varón	55	12.56	3.190	0.430					

Nota. Suma de los valores de los ítems de motivo superficial.

TABLA 4.110.

Estrategia superficial (ítems 32, 36, 40, 44 y 48) en cada género. Estadísticos.

sexo	N	Media	δ	Error T.	d	F	p	t	p
mujer	216	13.97	4.385	0.299	-0.04	1.267	>0.05	-0.275	>0.05
varón	55	14.15	3.856	0.520					

Nota. Suma de los valores de los ítems de estrategia superficial.

En cuanto a los resultados académicos, no hubo ninguna correlación entre el sexo y la nota obtenida en el examen (r de Pearson de -0.024 , no significativa), con ninguno de los dos tipos de test. Tampoco hubo correlación con la habilidad al manejar el test basado en confianza (*TBC-aciertos*, r de 0.014 , no significativa).

En el siguiente cuadro veremos los estadísticos básicos de los resultados en el examen (TBC, aciertos, TBC-aciertos) en cada uno de los dos grupos de estudiantes según el género:

TABLA 4.111.

Estadísticos de los resultados académicos separados por género.

sexo	N	Nota TBC	δ	Error T.	d	F	p	t	p
mujer	216	6.21	1.85	0.127	0.07	2.975	>0.05	0.483	>0.05
varón	55	6.07	2.21	0.298					
sexo	N	Nota por aciertos	δ	Error T.	d	F	p	t	p
mujer	216	6.27	1.81	0.124	0.10	1.398	>0.05	0.572	>0.05
varón	55	6.09	2.13	0.287					
sexo	N	TBC-aciertos	δ	Error T.	d	F	p	t	p
mujer	216	-0.04	0.584	0.040	-0.03	0.186	>0.05	-0.243	>0.05
varón	55	-0.02	0.622	0.084					

Como se puede apreciar, ambos géneros obtuvieron similares resultados en el examen, tanto corregido según aciertos como según el grado de confianza. La habilidad del estudiante en la utilización del TBC (TBC-aciertos), es igualmente similar. No hay diferencias significativas entre los géneros en estos parámetros de resultados académicos.

En definitiva, en nuestros grupos, los varones y las mujeres obtienen notas similares. El TBC no beneficia a ningún sexo, carece de sesgo de género. La única diferencia de cierta entidad entre los hombres y las mujeres en nuestra encuesta fue que estas últimas declaran mayor nivel de estrés antes de los exámenes.

4.4. ESTUDIO FACTORIAL DE LA ENCUESTA Y DE LAS ESCALAS DE ENFOQUES DE APRENDIZAJE Y AUTORREGULACIÓN

Hemos hecho el estudio factorial de la encuesta utilizando el programa SPSS en sus funciones de análisis factorial exploratorio, utilizando la rotación ortogonal VARIMAX para obtener los componentes factoriales. Al haberle realizado la encuesta a 272 estudiantes había número suficiente de sujetos para poder realizar este análisis (habitualmente se establece el mínimo necesario de sujetos en 200). Los datos detallados del estudio factorial exploratorio están en el anexo H.

4.4.1. ESCALA R- SPQ–2F (ENFOQUES DE APRENDIZAJE, DE BIGGS)

En primer lugar hicimos el análisis para los 20 ítems de la encuesta R-SPQ–2F, de enfoques de aprendizaje de Biggs. Era factible realizarlo al haber 13.5 sujetos por cada ítem, por encima de la relación mínima de 10.

El resultado de la *prueba de adecuación muestral KMO* (de Kaiser–Meyer–Olkin) ha sido de 0.811, muy por encima del mínimo exigido para hacer un análisis factorial que es de 0.500.

La *prueba de esfericidad de Bartlett*, que indica si hay correlaciones en la matriz, también ha dado resultados que permiten hacer el análisis factorial, al ser su significación $p < 0.000$ (es válida si $p < 0.05$).

La tabla de componentes del análisis factorial de la encuesta R-SPQ–2F, de Biggs (datos completos en el anexo H) arrojó que había 5 componentes principales con autovalores superiores a 1, que son los componentes o factores válidos. Los mostramos en la tabla 4.112, donde se exponen los valores únicamente de los 5 componentes.

TABLA 4.112.

Componentes válidos del análisis factorial de la encuesta.

KMO = 0.811; esfericidad de Bartlett 1345.304; $p < 0.0001$					
Varianza total explicada					
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	4.690	23.451	23.451	4.690	23.451
2	2.425	12.126	35.577	2.425	12.126
3	1.441	7.204	42.780	1.441	7.204
4	1.344	6.720	49.500	1.344	6.720
5	1.035	5.175	54.675	1.035	5.175

Como se ve en la tabla anterior, estos 5 factores explicaban una varianza total del 54.675 %, suficiente para el análisis factorial.

A continuación hicimos la rotación de componentes. Aplicamos la rotación ortogonal, o rotación VARIMAX, para factores no correlacionados entre sí o correlacionados de forma muy débil (tabla 4.113).

TABLA 4.113.

Matriz de componentes rotados, SPQ-2F.

Matriz de componentes rotados					
ÍTEMS	Componente				
	1	2	3	4	5
34	.700				
46	.658				
41	.658				
42	.646			-.372	
37	.641				
33	.563				
45	.554				
29	.533			.412	
38	.472		-.437		
47		.769			
44		.730		.309	
43		.605			
48		.507			
36			.758		
39			.698		
30	.422		-.547		
32				.747	
40		.326		.590	
35					.784
31				.307	.657

Nota. Datos detallados en anexo H.

Aparecen cinco agrupamientos bien definidos. El porcentaje de varianza explicada con estos cinco agrupamientos es de 54.675 %. Vamos a exponer las características de los cinco agrupamientos, siguiendo el orden de las columnas de izquierda a derecha.

4.4.1.1. PRIMER COMPONENTE

El primer agrupamiento (1ª columna por la izquierda) tiene 10 elementos, que son los ítems (en orden decreciente de correlación) 34, 46, 41, 42, 37, 33, 45, 29, 38 y 30.

Son todos los ítems de enfoques profundos de aprendizaje. Separándolos entre *motivos* y *estrategias*, la relación de ítems, ordenados, queda así:

TABLA 4.114

Tabla de correlaciones del primer componente.

Ítem 34, <i>ESTRATEGIA PROFUNDA</i>	.700
Ítem 46, <i>ESTRATEGIA PROFUNDA</i>	.658
Ítem 41, <i>MOTIVO PROFUNDO</i>	.658
Ítem 42, <i>ESTRATEGIA PROFUNDA</i>	.646
Ítem 37, <i>MOTIVO PROFUNDO</i>	.641
Ítem 33, <i>MOTIVO PROFUNDO</i>	.563
Ítem 45, <i>MOTIVO PROFUNDO</i>	.554
Ítem 29, <i>MOTIVO PROFUNDO</i>	.533
Ítem 38, <i>ESTRATEGIA PROFUNDA</i>	.472
Ítem 30, <i>ESTRATEGIA PROFUNDA</i>	.422

Podemos ver en la tabla 4.114 que el estudio factorial identifica perfectamente los enfoques profundos, pero no diferencia entre los «*motivos*» y las «*estrategias*», que aparecen intercalados en el listado ordenado. Nuestro análisis factorial no distingue entre estrategias y motivos profundos. No obstante, al correlacionar los enfoques profundos con los demás ítems de la encuesta sí aparecía nítida la distinción entre *profundos* y *superficiales*, como se puede ver en el capítulo 4.2.2 de «Resultados».

En nuestra encuesta, con nuestros alumnos, se agrupan los enfoques de aprendizaje profundos, pero la diferencia entre motivos y estrategias se difumina; se agrupan en un solo componente de 10 elementos (*motivos* y *estrategias*).

4.4.1.2. SEGUNDO COMPONENTE

El segundo componente –correspondiente a la segunda columna- es un factor compuesto de la agrupación de 5 ítems. Los 5 ítems son de *enfoques superficiales* de aprendizaje. La tabla con la lista de ítems es la siguiente:

TABLA 4.115.

Tabla de correlaciones internas del segundo componente.

Ítem 47, <i>MOTIVO SUPERFICIAL</i>	0.769
Ítem 44, <i>ESTRATEGIA SUPERFICIAL</i>	0.730
Ítem 43, <i>MOTIVO SUPERFICIAL</i>	0.605
Ítem 48, <i>ESTRATEGIA SUPERFICIAL</i>	0.507
Ítem 40, <i>ESTRATEGIA SUPERFICIAL</i>	0.326

Como vimos en el caso anterior, este agrupamiento incluye tanto *motivos* como *estrategias* superficiales; el análisis no distingue entre unos y otros. En este agrupamiento se incluyen la mitad de las preguntas que ponderan los enfoques superficiales, aunque el último ítem –el 40– se encuentra con mayor correlación en otro agrupamiento. Los enunciados de estos ítems son:

47) No le encuentro sentido a aprender contenidos que probablemente no caerán en el examen (0.769).

44) Creo que los profesores no deberían esperar que los alumnos dedicaran mucho tiempo a estudiar cosas que no van a caer en el examen (0.730)

43) Me parece que no ayuda estudiar los temas en profundidad. Confunde y hace perder el tiempo cuando todo lo que se necesita es un conocimiento por encima de los temas (0.605).

48) Me parece que la mejor manera de aprobar es recordar las respuestas de las preguntas que con más probabilidad pueden caer en el examen (0.507).

40) Generalmente limito mi estudio a lo que está específicamente ordenado, porque creo que es innecesario hacer más (0.326).

La lectura de los enunciados de estas preguntas explica su estrecha correlación; hacen referencia a las tareas a realizar para aprobar el examen –se menciona explícitamente el examen en tres de ellos, y está implícito en los otros dos– desde un enfoque superficial y de logro. Podríamos decir que el enunciado del agrupamiento es «aprobar los exámenes con el menor esfuerzo ».

El segundo componente, compuesto de ítems de enfoques superficiales, muestra un rasgo de logro dentro de un enfoque superficial (aprobar el examen con el mínimo esfuerzo)

4.4.1.3. TERCER COMPONENTE

El tercer agrupamiento consta de solamente 4 ítems. Aquí los exponemos en la tabla 4.116.

TABLA 4.116.

Tercer componente (ítems 36, 39, 30 y 38)

Ítem 36, <i>ESTRATEGIA SUPERFICIAL</i>	0.758
Ítem 39, <i>MOTIVO SUPERFICIAL</i>	0.698
Ítem 30, <i>ESTRATEGIA PROFUNDA</i>	-0.547
Ítem 38, <i>ESTRATEGIA PROFUNDA</i>	-0.437

Como vimos en la tabla 4.114, dos de ellos, el 30 y el 38, son de enfoques profundos –estrategia profunda– y tienen correlación negativa. Los otros dos, el 36 y el 39 son ítems de enfoques superficiales, estrategia el 36 y motivo el 39. Los enunciados de estos ítems son:

36) Aprendo las cosas repitiéndolas hasta que me las sé de memoria, aunque no las entienda (0.758).

39) Creo que puedo aprobar la mayoría de las evaluaciones memorizando los aspectos clave en lugar de intentar comprenderlos (0.698).

30) Al elaborar o estudiar un tema, no me encuentro satisfecho hasta que me he formado mis propias conclusiones sobre él (–0.547).

38) Me hago preguntas a mí mismo sobre los temas importantes hasta que los comprendo completamente (–0.437).

En el enunciado de las preguntas se ve que los dos primeros ítems –de enfoque superficial– hacen referencia a la memorización como método de aprendizaje, mientras que los dos últimos –enfoques profundos con correlación negativa– hacen referencia a la comprensión –«me he formado mis propias conclusiones», «los comprendo perfectamente»– como procedimientos de aprendizaje. Este componente pone de manifiesto una clara oposición entre el aprendizaje memorístico y el aprendizaje basado en la comprensión y la interiorización de los conceptos.

Este factor, el *tercer componente*, está relacionado con la percepción por el alumno de una oposición entre la memorización y la comprensión a la hora de estudiar.

4.4.1.4. CUARTO COMPONENTE

El cuarto agrupamiento consta de 6 ítems, pero solo los dos primeros son exclusivos de este componente, mientras que los otros 4 se integran, con mayor intensidad, en otros componentes. Aquí los exponemos en el siguiente cuadro.

TABLA 4.117.

Cuadro de correlaciones internas del cuarto componente.

Ítem 32, <i>ESTRATEGIA SUPERFICIAL</i>	0.747	
Ítem 40, <i>ESTRATEGIA SUPERFICIAL</i>	0.590	
Ítem 29, <i>MOTIVO PROFUNDO</i>	0.412	Componente 1
Ítem 44, <i>ESTRATEGIA SUPERFICIAL</i>	0.309	Componente 2
Ítem 31, <i>MOTIVO SUPERFICIAL</i>	0.307	Componente 5
Ítem 42, <i>ESTRATEGIA PROFUNDA</i>	-0.372	Componente 1

Nota. Datos detallados en el anexo H.

Hemos puesto en el recuadro de la derecha los agrupamientos en los que cada uno de los 4 ítems que pertenecen a otros agrupamientos se integra.

Llama poderosamente la atención la inclusión en este agrupamiento, con una correlación positiva de 0.412, de un ítem de motivo profundo –el 29–. No resulta fácil encontrar explicación para esto, salvo por el hecho de que este ítem habla sobre el estudio, al igual que los demás; los demás para expresar que el alumno limita su estudio a lo imprescindible y el ítem 29 para expresar que el alumno siente satisfacción personal al estudiar; de alguna manera los alumnos no consideran ese agrupamiento contradictorio.

El enunciado de los ítems de este agrupamiento es:

32) Sólo estudio seriamente lo que se da en las clases o lo que está en los programas detallados de las asignaturas. **0.747**

40) Generalmente limito mi estudio a lo que está específicamente ordenado, porque creo que es innecesario hacer más. **0.590**

29) Me doy cuenta de que estudiar me proporciona con frecuencia un sentimiento de profunda satisfacción personal. **0.412**. *Se correlaciona más con el factor 1.*

44) Creo que los profesores no deberían esperar que los alumnos dedicaran mucho tiempo a estudiar cosas que no van a caer en el examen. **0.309**. *Se correlaciona más con el factor 2.*

31) Mi objetivo es aprobar el curso haciendo el mínimo trabajo posible. **0.307**. *Se correlaciona mucho más con el factor 5.*

42) Empleo bastante de mi tiempo libre en buscar más información sobre temas interesantes que se han discutido en las diferentes clases. **0.372**. *Se correlaciona más, y en positivo, con el factor 1.*

Los ítems de este agrupamiento, con excepción del mencionado 29, expresan que el alumno hace un esfuerzo mínimo a la hora de estudiar. A diferencia del componente 2, aquí no hay referencias a los resultados –el logro de un aprobado–, salvo en el ítem 31 que es el más débil de este componente; el alumno, simplemente, limita su esfuerzo de estudio con independencia de que le llegue o no para conseguir aprobar. El último ítem de la lista –el 42– que se correlaciona negativamente con este agrupamiento, expresa la idea de dedicar más tiempo del estrictamente necesario al estudio ampliando la información en los ratos libres, justo lo opuesto que expresan los demás ítems del componente 4.

Este agrupamiento, el cuarto componente, expresa la tendencia a hacer el «mínimo esfuerzo» para estudiar.

4.4.1.5. QUINTO COMPONENTE

El quinto agrupamiento consta solo de 2 ítems; aquí los exponemos en la tabla 4.118.

TABLA 4.118.

Cuadro de correlaciones internas del quinto componente.

Ítem 35, <i>MOTIVO SUPERFICIAL</i>	0.784	
Ítem 31, <i>MOTIVO SUPERFICIAL</i>	0.657	

Nota. Datos detallados en el anexo H.

El enunciado de estos ítems es el siguiente:

35) Cuando veo que la asignatura no me resulta muy interesante, trabajo lo menos posible, *correlación 0.784*.

31) Mi objetivo es aprobar el curso haciendo el mínimo trabajo posible, *correlación 0.657, correlación 0.657*.

La semejanza de ambos ítems es evidente al leer sus enunciados; ambos hacen referencia a actitudes y en ambos se menciona «trabajar lo menos posible o mínimo trabajo posible». Expresan un concepto que podríamos definir como «*mínimo trabajo necesario*». En el ítem 35 existe el matiz de limitar esta falta de esfuerzo a unas cuantas asignaturas –«las que no me resultan interesantes»–, mientras que en el 31 esta actitud de mínimo esfuerzo se extiende, en general, a todas las asignaturas aunque con la condición de «aprobar el curso».

Este quinto componente expresa una actitud muy parecida a la del 4º componente –que era la de *estudiar lo mínimo*–, de hecho, el ítem 31 se correlaciona, aunque menos, con el componente 4. No resulta fácil explicar por qué estas dos preguntas funcionan

como un agrupamiento semi-aislado; tal vez la no mención al estudio en concreto, sino al término más genérico «trabajo» esté en la raíz de ello.

Lo común a estos dos ítems tan correlacionados, el 31 y el 35, es que reflejan una postura de trabajar lo menos posible, al igual que el agrupamiento anterior, de «tendencia al esfuerzo mínimo»

4.4.2. ANÁLISIS FACTORIAL DE LA ENCUESTA R-SPQ-2F CON SOLO DOS AGRUPAMIENTOS

Haciendo el análisis confirmatorio de los resultados de la encuesta de Biggs para solamente dos componentes, obtenemos la tabla 4.119.

TABLA 4.119.

Agrupamientos y correlaciones de la encuesta RSPQ-2F con dos componentes.

ÍTEMS	Matriz de componentes rotados ^a		esfericidad de Bartlett = 1345.304; p<0.001 KMO = 0.811;
	1	2	
34	.706		Estrategia profunda
42	.676		Estrategia profunda
46	.672		Estrategia profunda
41	.660		Motivo profundo
37	.640		Motivo profundo
45	.575		Motivo profundo
33	.533		Motivo profundo
29	.508		Motivo profundo
38	.500		Estrategia profunda
30	.440		Estrategia profunda
36		.657	<i>Estrategia superficial</i>
48		.651	<i>Estrategia superficial</i>
44		.610	<i>Estrategia superficial</i>
47		.598	<i>Motivo superficial</i>
40		.591	<i>Estrategia superficial</i>
39		.561	<i>Motivo superficial</i>
43		.534	<i>Motivo superficial</i>
31		.501	<i>Motivo superficial</i>
32		.416	<i>Estrategia superficial</i>
35		.396	<i>Motivo superficial</i>

Nota. Datos detallados en el anexo H.

Como se puede ver en la tabla, el análisis factorial con 2 componentes y rotación VARIMAX agrupa los enfoques profundos en un solo componente y los superficiales en el otro, pero no separa nítidamente los ítems de *motivo* de los de *estrategia* en ninguno de los dos casos. El porcentaje total de varianza explicada con solo dos componentes, fue del 35.58 %, un porcentaje reducido, que denota que considerando solo los factores «enfoque profundo/superficial» queda la mayoría de la varianza sin explicar. Cuando hicimos el análisis exploratorio, vimos que con 5 componentes el porcentaje de varianza explicada se elevaba al 54.65 %.

El análisis factorial establece una clara separación entre enfoques profundos y superficiales, pero en nuestro análisis no encontramos que haya una separación nítida entre motivos y estrategias en cada componente, superficial o profundo.

4.4.3. ANÁLISIS FACTORIAL DE LAS PREGUNTAS SOBRE *AUTORREGULACIÓN*

Hemos sometido también los cinco ítems sobre autorregulación al análisis factorial mediante el programa SPSS. Los resultados han sido:

El resultado de la *prueba de adecuación muestral KMO* (de Kaiser–Meyer–Olkin) ha sido de 0.750, muy por encima del mínimo exigido de 0.500.

La *prueba de esfericidad de Bartlett* también ha dado resultados que permiten hacer el análisis factorial, al ser su significación $p < 0.0001$ (es válida si $p < 0.05$).

TABLA 4.120.

Análisis factorial de los ítems de autorregulación.

Varianza total explicada					
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	2.542	50.838	50.838	2.542	50.838
2	.861	17.222	68.060		
3	.707	14.138	82.198		
4	.460	9.193	91.391		
5	.430	8.609	100		

Como se ve en la tabla, el primer componente satura por si solo más del 50.8 % de la varianza, por lo que en solo hay un agrupamiento, que en la tabla 4.121 se ve que corresponde al de los 5 ítems de autorregulación.

TABLA 4.121.

Ítems del componente «autorregulación»

Matriz de componentes	
	Componente 1
Ítem 51	0.769
Ítem 53	0.747
Ítem 52	0.729
Ítem 49	0.721
Ítem 50	0.586

En la tabla 4.121 se aprecia que los ítems 51, 53, 52 y 49 tienen una alta correlación con el agrupamiento, bastante homogénea en todos ellos. En cambio, el ítem 50 tiene una correlación sensiblemente inferior, aunque elevada.

Este agrupamiento tiene una consistencia interna (medida con el índice de Cronbach) de $\alpha = 0.735$, como vemos en la tabla 4.122. Consideramos que es una consistencia interna elevada teniendo en cuenta el número reducido de ítems que tiene este factor.

TABLA 4.122.
Estadísticos de fiabilidad

Estadísticos de fiabilidad	
Alfa de Cronbach	N° de elementos
0.735	5

En la tabla 4.123, donde se expone la aportación de cada ítem a la consistencia interna se puede apreciar que el ítem 50 disminuye la consistencia interna, que sería mayor ($\alpha = 0.751$ en lugar de 0.735) si ese ítem se eliminase.

TABLA 4.123.

Tabla de estadísticos total–elemento

Estadísticos total–elemento				
	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento–total corregida	Alfa de Cronbach si se elimina el elemento
Ítem 49	17.97	10.940	.503	.689
Ítem 50	19.15	9.469	.404	.751
Ítem 51	18.43	9.899	.578	.657
Ítem 52	18.01	10.590	.545	.673
Ítem 53	17.72	11.226	.537	.682

A pesar de ese elemento más débil, el ítem 50, la escala de autorregulación tiene una adecuada consistencia interna y constituye un factor bien definido, por lo que decidimos mantenerla incluso con ese ítem.

4.4.4. ANÁLISIS FACTORIAL DE LOS DEMÁS ÍTEMS DE LA ENCUESTA

Los parámetros que ponderan la adecuación del estudio factorial dan resultados favorables, con la medida KMO, de adecuación muestral, de 0.700, y la prueba de esfericidad de Bartlett con significación $p < 0.0001$. Estos datos se exponen en la tabla 4.124.

TABLA 4.124.
Medidas de adecuación muestral, ítems 2–28

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser–Meyer–Olkin.		.700
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2934.064
	gl	351
	Sig.	.000

El estudio factorial con rotación ortogonal da la matriz de componentes rotados que exponemos en la tabla 4.125; en ella se obtienen 9 componentes que saturan la varianza en un 70.28 %.

TABLA 4.125.
Matriz de componentes rotados; ítems 2–28.

Matriz de componentes rotados^a								
Ítem	Componente							
	1	2	3	4	5	6	7	8
20	.804							
23	.795							
10	.792							
17	.759							
7	.511							-.351
8		.807						
21		.785						
15		.702						
2		.457			-.381			
27			.916					
26			.882					
28			.845					
9		.312		.717				
16				.696				-.318
22				.665	.358			
6				.618				
24					.817			
3					.768			
25					.564			.508
11						.808		
13						-.674		
12	.344			.425		.593		
19				.451			.721	
18		.552					.712	
4								.753
14								
5		.450						

Continuación (matriz de componentes rotados)	
	Componente
	9
Ítem 6	.398
Ítem 14	-.736
Ítem 5	.619

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser. A

a. La rotación ha convergido en 11 iteraciones.

Nota. Datos detallados en el anexo H.

4.4.4.1. COMPONENTES QUE MIDEN LA «EFICACIA DOCENTE» DE CADA TIPO DE EXAMEN

De los 9 componentes que el análisis factorial de la encuesta nos proporciona, 4 de ellos, el 1º, 2º y 4º, hacen referencia a los tres agrupamientos que de forma intuitiva nosotros habíamos establecido en el epígrafe 4.2.1.3 y que medía lo que llamamos «eficacia docente de cada examen». Los tres agrupamientos que habíamos formado intuitivamente se expresaban en la tabla 4.48, que reproducimos a continuación.

TABLA 4.48.

Agrupamientos con cada tipo de examen.

Opinión del alumno sobre el <i>test basado en aciertos</i>	Ítems 5, 8, 15, 18 y 21
Opinión del alumno sobre el TBC	Ítems 6, 9, 16, 19 y 22
Opinión del alumno sobre el examen de <i>preguntas abiertas</i>	Ítems 7, 10, 17, 20 y 23

Como vamos a exponer, estos tres factores se confirman en el examen factorial, correspondiendo al componente 2 (*test basado en aciertos*), al 4 (TBC) y al 1 (examen de *preguntas abiertas*). Los resultados de este análisis factorial ya los expusimos, resumidos, en las tablas 4.50, 4.51 y 4.52.

El componente 1 está formado por los siguientes ítems y ofrece los siguientes resultados:

TABLA 4.126.

Tabla de correlaciones del componente 1, correspondiente al examen de tipo *preguntas abiertas*.

FACTOR 1, matriz de correlaciones de los ítems que lo componen.	
Ítem 20, el examen de preguntas abiertas te hace darte cuenta de tus <u>lagunas formativas</u> .	0.804
Ítem 23, el examen de preguntas abiertas te hace aprender más	0.795
Ítem 10, el examen de preguntas abiertas examen te obliga a <u>entender bien todo</u>	0.792
Ítem 17, el examen de preguntas abiertas es el que permite al profesor conocer mejor	0.759
Ítem 7, el examen de preguntas abiertas te obliga dedicar más tiempo al estudio	0.511

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este agrupamiento hace referencia a las cualidades percibidas por el alumno del examen de preguntas abiertas; *eficacia docente con el examen de tipo preguntas abiertas*.

El factor 2 aparece expuesto aisladamente en la tabla 4.127.

TABLA 4.127.

Tabla de correlaciones del componente 2, que corresponde al examen de tipo *test por aciertos*.

FACTOR 2, matriz de correlaciones de los ítems que lo componen.	
Ítem 8, el examen de test por aciertos te obliga a entender bien todo.	0.807
Ítem 21, el examen de test por aciertos te hace aprender más	0.785
Ítem 15, el examen de test por aciertos es el que permite al profesor conocer mejor...	0.702
Ítem 18, el examen de test por aciertos te hace darte cuenta de tus lagunas formativas	0.552
<i>Ítem 2, el examen de test por aciertos es el que más te gusta</i>	0.457
Ítem 5, el examen de test por aciertos te obliga dedicar más tiempo al estudio	0.450

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este agrupamiento hace referencia a las cualidades percibidas por el alumno del examen de test corregido por número de aciertos; *eficacia docente con el examen de tipo test por aciertos*. Aparece un ítem no homólogo a los que aparecen en la tabla anterior y es el ítem 2, referido a la preferencia del alumno por el examen de tipo test por aciertos (en letra bastardilla). Los demás son los homólogos a los que aparecen en la tabla 4.126.

A continuación incluimos el cuarto factor, que es el homólogo de los dos anteriores, dejando para después el tercer agrupamiento, que explora un aspecto muy diferente. El cuadro de correlaciones internas del cuarto factor se expone en la tabla 4.128.

TABLA 4.128.

Correlaciones del componente 1, examen de tipo *test basado en la confianza* (TBC).

FACTOR4, matriz de correlaciones de los ítems que lo componen.	
Ítem 9, el TBC te hace entender bien todo...	0.717
Ítem 16, el TBC es el que permite al profesor conocer mejor...	0.696
Ítem 22, el TBC te hace aprender más	0.665
Ítem 6, el TBC te obliga a dedicar más tiempo al estudio	0.618
Ítem 19, el TBC te hace darte cuenta de tus lagunas formativas	0.451
Ítem 12, el TBC te obliga a aprender más datos de memoria	0.425

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este agrupamiento hace referencia a las cualidades percibidas por el alumno del examen de test basado en la confianza; *eficacia docente con el TBC*. Aparece un ítem no homólogo a los que aparecen en la tabla anterior y es el ítem 12, referido a la exigencia de este examen a dedicar más tiempo al estudio; ese ítem, el 12, en realidad se correlaciona más con otro componente distinto, con el agrupamiento 6. Los demás son los homólogos a los que aparecieron en la tabla 4.126.

Estos tres componentes que hemos visto corresponden a los tres agrupamientos que de nosotros habíamos establecido en el capítulo 4.2.1.3 y en la tabla 4.48. Y que llamábamos «eficacia docente» del examen de *preguntas abiertas*, del examen de *test por aciertos* y del TBC. Como hemos visto, en el caso del test por aciertos y del TBC en estos agrupamientos hay un elemento más, el ítem 2 en el primer caso y el 12 en el segundo, ambos, sin embargo, se correlacionan más con otros factores. Hemos comprobado que los tres agrupamientos establecidos por nosotros se confirman en el análisis factorial exploratorio. La consistencia interna de cada uno de los agrupamientos se muestra en la tabla 4.129.

TABLA 4.129.

Fiabilidad de los agrupamientos «eficacia docente» de cada tipo de examen.

AGRUPAMIENTO	α de Cronbach
Factor «eficacia docente del test por aciertos» (2)	0.785
Factor «eficacia docente del TBC» (4)	0.737
Factor «eficacia docente de las preguntas abiertas» (1)	0.891

Nota. Resumido de los datos del capítulo 4.2.1.3

4.4.4.2. OTROS AGRUPAMIENTOS QUE SE OBTIENEN DEL ANÁLISIS FACTORIAL

Los otros componentes que aparecen en el análisis de agrupamientos se muestran a continuación.

Factor 3

El agrupamiento 3 está constituido por tres ítems, el 26, 27 y 28, que miden el nivel de estrés autopercebido por el estudiante al enfrentarse ante un tipo de examen. Las correlaciones internas de este agrupamiento se muestran en la tabla 4.130.

TABLA 4.130.

Tabla de correlaciones internas del componente 3.

FACTOR3, matriz de correlaciones de los ítems que lo componen.	
Ítem 27, te sientes estresado al tener que hacer el TBC	0.916
Ítem 26, te sientes estresado al tener que hacer el test por aciertos	0.882
Ítem 28, te sientes estresado al tener que hacer el examen de preguntas abiertas	0.845

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este factor mide el estrés que siente el alumno al hacer cada tipo de examen, y la alta correlación de los ítems indica que los alumnos que reconocen estresarse con un tipo de examen se estresan también con los demás. A este agrupamiento podríamos llamarlo «estrés autopercebido ante los exámenes».

Factor 5

El factor 5 se compone de cuatro elementos, uno de los cuales, el correspondiente al ítem 22, está incluido con una mayor correlación en el factor 4.

TABLA 4.131.

Tabla de correlaciones internas del componente 5.

FACTOR 5, matriz de correlaciones de los ítems que lo componen.	
Ítem 24, en los TBC, suelo escoger bien el grado de seguridad	0.817
Ítem 3, me gusta el TBC	0.768
Ítem 25, este curso espero tener varios sobresalientes	0.564
Ítem 22, el TBC me hace aprender más	0.358

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este factor agrupa el desenvolverse bien con el TBC (ítem 24), el preferir ese sistema de examen (ítem 3), la expectativas positivas de resultados en el aprendizaje (ítem 25) y, en menor medida, la sensación de que el TBC ayuda a aprender (ítem 22). Podríamos definirlo como *«habilidad en el uso del TBC, gusto por él y autoeficacia docente»*.

Factor 6

Este agrupamiento se compone de tres elementos, que hacen referencia al esfuerzo de memorización a que obliga cada examen. Lo destacado de este agrupamiento es que uno de los ítems se correlaciona intensamente con los demás, pero con correlación negativa. Las cifras de esta correlación las exponemos en la tabla 4.132.

TABLA 4.132.

Tabla de correlaciones internas del componente 6.

Ítem 11, el test por aciertos te obliga a aprender más datos de memoria.	0.808
Ítem 13, las <i>preguntas abiertas</i> te obligan a aprender más datos de memoria.	-0.674
Ítem 12, el TBC te obliga a aprender más datos de memoria.	0.593

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este factor refleja la idea de que los hay una oposición entre las ideas de que los exámenes de tipo *test-por aciertos* y TBC–obligan memorizar mucho y la de que le examen de preguntas abiertas obliga a memorizar. Los alumnos que opinan que un tipo cualquiera de examen de test obliga a hacer un esfuerzo de memorización grande tienden a minusvalorar el esfuerzo de memorización a que obliga el examen de preguntas abiertas.

Factor 7

Este factor se compone solo de dos ítems, ambos correspondientes a la pregunta que valora la influencia de cada examen a la autoevaluación del alumno -«darte cuenta de tus lagunas formativas»-. Exponemos las correlaciones internas en la tabla 4.133.

TABLA 4.133.

Tabla de correlaciones internas del componente 7

Ítem 19, el TBC te hace darte cuenta de tus <u>lagunas formativas</u>	0.721
Ítem 18, el <i>test por aciertos</i> te hace darte cuenta de tus <u>lagunas formativas</u>	0.712

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este factor expresa la opinión de que los test son buenos procedimientos para detectar lagunas formativas, tanto sean *test por aciertos* como TBC. Ambos ítems, el 19 y el 18, han sido incluidos ya en los agrupamientos 2 y 4, respectivamente.

Factor 8

El factor 8 está integrado por dos ítems –el 4 y el 25- con correlación positiva y otros dos, el 7 y el 16, con una correlación de signo negativo, y de menor cuantía. Los dos ítems de correlación positiva hacen referencia a preferencia por el examen de tipo *preguntas abiertas* y la

pregunta sobre la autoeficacia docente. Las correlaciones internas de este agrupamiento las mostramos en la tabla 4.134.

TABLA 4.134.

Tabla de correlaciones internas del componente 8.

Ítem 4, me gusta el examen de preguntas abiertas	0.753
Ítem 25, creo que este curso tendré varios sobresalientes	0.508
Ítem 7, el examen de preguntas abiertas te obliga dedicar más tiempo al estudio	-0.351
Ítem 16, el TBC le permite al profesor conocer mejor lo que el alumno sabe	-0.318

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

En este agrupamiento se ve que la opinión favorable al examen de *preguntas abiertas* –ítem4–, el que los alumnos consideran más exigente en cuanto al tiempo que hay que dedicar a prepararlo, se asocia a buenas expectativas académicas –ítem 25–. También se observa que este grupo de alumnos, con preferencia por las *preguntas abiertas*, opinan que ese tipo de examen obliga a dedicar menos tiempo al estudio que los demás y no tienen una opinión tan favorable como los demás respecto a que el TBC es un buen sistema de evaluación –ítem 16–.

Factor 9

Los ítems y el cuadro de correlaciones internas del factor 9 los exponemos en la tabla 4.135.

TABLA 4.135.

Tabla de correlaciones internas del componente 9.

Ítem 14, el tipo de examen esperado influye en mi forma de estudiar	-0.736
Ítem 5, el test por aciertos me obliga a dedicar más tiempo al estudio	0.619
Ítem 6, el TBC me obliga a dedicar más tiempo al estudio	0.398

Nota. Datos detallados en el anexo H. Todas las correlaciones son significativas, $p < 0.0001$

Este último componente expresa una idea muy interesante: los alumnos que declaran dedicar mucho tiempo al estudio cuando se están preparando un examen de tipo test, –sea test *por aciertos* o TBC, aunque mucho más en el primer caso– son los que opinan que el tipo de examen que esperan no influye –o influye menos– en su forma de estudiar. Dando la vuelta al argumento, esto significaría que los alumnos que opinan que su forma de estudiar se ve muy influida por el tipo de examen que esperan, son los que reconocen que el examen de tipo *preguntas abiertas* es el que les obliga a dedicar más tiempo al estudio. Los alumnos en los que el tipo de examen esperado no influye en su estudio, dedican más tiempo al estudio aunque esperen un examen de tipo test por aciertos, que es visto como el menos exigente a este respecto. Es una conclusión que parece absolutamente lógica.

4.5. ESTUDIO DE LOS PARÁMETROS QUE MIDEN LA AUTOCONFIANZA MEDIA DEL ALUMNO

4.5.1. ESTUDIO DE LA *PUNTUACIÓN EN CONFIANZA (PC)* Y DEL *ÍNDICE DE CONFIANZA (IC)*

Como hemos explicado en Metodología, La *Puntuación en confianza* (acrónimo *PC*) es el parámetro que mide la elección media del alumno en cuanto al grado de confianza y el *índice de confianza (IC)* es el cociente entre la *PC* y la nota por aciertos, que nos indica la tendencia a asumir riesgos en función de sus conocimientos.

La PC utilizaba la siguiente fórmula [(*número de opciones de máxima seguridad*) \times 10 + (*número de opciones de seguridad intermedia*) \times 5]/ *número de preguntas*.

El IC se obtenía así: *PC / nota por aciertos*.

4.5.1.1. CORRELACIÓN ENTRE LA PC Y EL RESULTADO ACADÉMICO

La correlación entre la PC y el resultado del examen nos indicará si las expectativas del alumno son realistas, hasta qué punto su nivel de confianza está fundamentado. Una baja correlación indicaría que la seguridad en acertar la pregunta que el alumno manifiesta no tiene una base real. Una correlación alta nos indicaría, en cambio, que el alumno domina su campo cognitivo y puede, en consecuencia, prever el resultado de la prueba. En el trabajo preliminar que hicimos con los alumnos de 2008 y 2009, calculamos la correlación entre PC y aciertos, análisis que repetimos en 2014:

TABLA 4.136.

Correlación entre *Nota según aciertos* y PC.

Año del estudio	Número de participantes	R de Pearson	Significación
2008, 2009	215 alumnos	0.704	p < 0.00001
2014	139 alumnos	0.681	p < 0.00001

Son correlaciones muy altas. Esto confirma los resultados que expusimos al principio del presente capítulo, que mostraban el porcentaje de aciertos medio con cada nivel de seguridad y que nos indicaba que la elección del grado de seguridad por los alumnos fue racional (tabla 4.3). Hicimos la correlación entre la PC y el número de

aciertos y no entre la PC y el resultado en el TBC porque en la puntuación obtenida en el TBC la confianza es un componente y no son, por lo tanto, variables independientes.

Otros autores, en el estudio internacional sobre los resultados PISA que hemos mencionado el Marco Teórico, han encontrado también una alta correlación entre la confianza del estudiante, medida durante el examen y referida a la probabilidad de acertar o fallar cada cuestión, y el resultado real de la evaluación (Morony, Kleitman, Lee y otros, 2013); en ese caso la correlación $-r$ de Pearson- entre ambos factores, confianza y resultado de la evaluación, fue de 0.60; en nuestro caso, como hemos visto, la correlación es de 0.70 en el estudio de 2008-2009 y de 0.68 en el de 2014.

4.5.1.2. RESULTADOS OBTENIDOS EN LA PC, ÍNDICE DE CONFIANZA (IC) Y PC-ACIERTOS

En estudio preliminar que hicimos en los estudiantes de 2008 y 2009, encontramos una estrecha correspondencia entre la puntuación en confianza y la nota obtenida en el *test por aciertos*. En la tabla 4.137 exponemos la IC en cada grupo de alumnos examinados.

TABLA 4.137.

IC de cada grupo de estudiantes encontrado en 2008 y 2009.

EXAMEN	1	2	3	4	5	Media
GRUPO	AMQ II 2008	AMQ II 2009	EMQ 2009	AMQ II 2009	AMQ I 2009	Total (5 exámenes)
I.C. (índice de confianza)	1.06	0.97	1.01	0.95	1.07	1.01

Nota. Los cinco grupos suman un total de 215 participantes.

El índice de confianza medio de los 215 alumnos fue de 1.01; es decir, la nota obtenida por número de aciertos era casi igual a la *puntuación en confianza*, como hemos indicado más arriba. Hay que destacar que mientras la primera nota tiene en cuenta los aciertos y los fallos, y tiene una penalización de 0.5 por cada uno de los últimos, la *puntuación en confianza* solo tiene en cuenta el nivel de seguridad escogido en cada pregunta, con independencia de que la respuesta fuese correcta o incorrecta.

En el trabajo posterior, aunque nuestro objetivo fundamental estaba en la encuesta, decidimos repetir las medidas de confianza en varios grupos de alumnos, para comprobar si se mantenían esos valores casi idénticos de la nota *por aciertos* y los parámetros para medir la confianza. Dispusimos de los datos para calcular la *PC* y el *IC* en dos grupos de alumnos, los correspondientes a los cursos de Enfermería y Fisioterapia de 2013–2014, en total 139 alumnos, 31 varones y 108 mujeres. Sobre ese total de 139 alumnos realizamos el presente análisis de datos. Los datos obtenidos se exponen en la tabla 4.138.

TABLA 4.138.
Estadísticos de la *puntuación en confianza* y del *índice de confianza*.

	Nota en TBC	Nota <i>por aciertos</i>	<i>PC</i>	<i>PC-aciertos</i> (<i>PC-NA</i>)	<i>IC</i>
Media	6.16	6.20	6.11	-0.09	1.08
Desviación t.	1.92	1.92	1.65	1.45	0.57
Error típico	0.188	0.180	0.156	0.091	0.056
Mediana	6.50	6.50	6.25	-0.08	0.98
F = 0.455;			p > 0.05 ;		Eta ² = 0.003

Nota. Datos detallados en el anexo F.

Al igual que vimos en el trabajo preliminar, al principio de este capítulo (tabla 4.2), la *Puntuación en confianza* es muy parecida a la nota real obtenida por los alumnos (en el TBC o *por aciertos*) y, en consecuencia, el *IC* está muy próximo a la unidad (1.08). Hay una pequeña diferencia en el *IC* entre 2008-2009 y 2013-2014, el *IC*

aumentó de 1.01 a 1.08. Este ligero aumento lo interpretamos como debido a la modificación en la tabla de puntuación que introdujimos en 2012, y que expusimos en el capítulo del marco conceptual, en las tablas 2.17 y 2.18; en la modificación que hicimos, guiados por el deseo de contrarrestar una tendencia a la excesiva prudencia por parte de algunos estudiantes, rebajamos la penalización por respuesta incorrecta con total seguridad de -4 puntos a -3 puntos, lo que creemos que pudo inducir a algunos alumnos a escoger opciones de mayor confianza, resultando en el mencionado aumento de IC –aumento pequeño, por otro lado-. La comparación de la PC y de la *nota por aciertos*, nos da lo que los investigadores australianos denominan «sesgo de puntuación» o «sobreconfianza» (Morony y otros, 2013; Stankov y otros, 2012). La gran proximidad de la media de ambos valores (PC y *nota por aciertos*) nos confirma que la ausencia de *sesgo de puntuación* –la situación en la que los conocimientos del alumno se corresponden exactamente con su confianza- se encuentra cuando la PC-NA es próxima a 0; los valores positivos indicarían «sobreconfianza» y los negativos «infraconfianza». Si considerásemos el sesgo mayor de 1 desviación estándar como el límite para considerar sobre confianza o infraconfianza significativa, los alumnos con PC-NA superior a 1.36 tendrían exceso de confianza y aquellos que lo tuvieran inferior a -1.54 tendrían falta de autoconfianza. Con el IC estas situaciones corresponderían a tener el IC mayor o menor de 1.

4.5.2. RELACIÓN ENTRE LA AUTOCONFIANZA Y LOS RESULTADOS DEL EXAMEN

Los parámetros *IC* y *PC–aciertos* sirven para medir la relación en cada alumno entre sus conocimientos reales y su confianza en ellos. Calculamos también la

diferencia, en valor absoluto, de la *puntuación en confianza* y la *nota por aciertos* en cada alumno, para evaluar en qué medida la disparidad entre la confianza del alumno y la nota obtenida influían en los resultados del examen. Este nuevo parámetro lo expresamos así: $\text{Puntuación en confianza} - \text{número de aciertos}$ o $|PC - NA|$.

Obtuvimos con estos índices los estadísticos básicos que mostramos en la tabla 4.139.

TABLA 4.139.

Estadísticos de la *puntuación en confianza* y del *índice de confianza*.

	IC	<i>PC-aciertos</i>	$ PC - NA $
Media	1.08	-0.09	1.11
Error típico	0.05	0.12	0.08
Desviación típica	0.572	1.446	0.928
Mediana	0.98	-0.08	0.87
Rango	5.38	8.93	5.35
Mínimo	0.59	-3.58	0.00
Máximo	5.97	5.35	5.35

Nota. Estudio realizado en una población de 139 estudiantes.

Hicimos la correlación de los resultados obtenidos por cada alumno en los índices PC e IC y los ítems del resto de la encuesta (utilizando el coeficiente r de Pearson). La primera cuestión en la que nos fijamos fue la relación del nivel de confianza (medido mediante la PC, el IC y el índice *PC-aciertos* y $|PC - NA|$) con la nota obtenida en el examen, tanto mediante el TBC como mediante la *puntuación por aciertos*, así como la diferencia de ambas. Los resultados obtenidos los exponemos en la tabla 4.140.

TABLA 4.140.

Tabla de correlaciones entre las medidas de la autoconfianza y las notas del examen.

	PC	IC	<i>PC-aciertos</i>	$ PC - NA $
Nota TBC	0.749 $p < 0.0001$	-0.471 $p < 0.0001$	-0.404 $p < 0.0001$	-0.215 $p < 0.05$
Nota por aciertos	0.681 $p < 0.0001$	-0.549 $p < 0.0001$	-0.549 $p < 0.0001$	-0.086 NS
TBC-aciertos	0.216 $p < 0.05$	0.236 $p < 0.01$	0.447 $p < 0.0001$	-0.401 $p < 0.0001$

Nota. Se utiliza la r de Pearson. Estudio realizado en 139 estudiantes.

En la tabla 4.140 se aprecia que hay una muy fuerte correlación entre el nivel de confianza y la nota obtenida por los dos procedimientos, como parece natural que sea. En el trabajo preliminar, realizado en 2008 y 2009, obtuvimos también una correlación negativa entre la nota y los parámetros que miden la confianza, en ese caso la correlación entre el IC y la nota por aciertos fue de -0.540, muy similar a la encontrada en el trabajo actual (-0.549). Otros autores han constatado este mismo hecho (Hench, 2012; Morony y otros, 2013).

Cuando se analiza el IC, se ve que tiene una correlación negativa con la nota obtenida por el alumno, lo que indica que **los alumnos peor preparados se arriesgan proporcionalmente más**. La correlación de 0.216 entre el parámetro TBC–aciertos, que mide la habilidad del alumno en el manejo del examen, y el PC, que mide la confianza, significa que los alumnos con mayor confianza son los que tienden a sacar más partido al TBC. Eso puede significar, también, que en nuestro sistema de puntuación el riesgo está, en realidad, poco penalizado. La correlación negativa alta y significativa (r de -0.401) entre el cociente en valor absoluto «confianza–aciertos» ($|PC-NA|$) y el cociente TBC–aciertos, significa que cuanto más parecida es la puntuación en confianza (PC) del alumno a la nota obtenida en el examen (siempre según el *número de aciertos*) tanto mayor puntuación recibe el alumno en el TBC, o dicho de otra manera, **el TBC premia a los alumnos que saben reconocer su auténtico nivel de conocimientos** y tienen una auto–confianza acorde a él, ni demasiado alta, ni demasiado baja.

En el TBC el alumno sale beneficiado cuando su nivel de autoconfianza es similar a su nivel de conocimientos. Cuando, por falta de sinceridad o déficit de metacognición, el alumno tiene muchas disparidad entre su nivel en autoconfianza y su número de aciertos, es penalizado por el sistema.

4.5.3. RELACIÓN ENTRE LA AUTOCONFIANZA Y EL GÉNERO DE LOS ESTUDIANTES

Otras correlaciones que analizamos fueron entre el nivel de confianza, el índice de riesgo, el cociente confianza–aciertos y el género del alumno, para responder a la cuestión: *¿influye el género del alumno en su asunción de riesgos?*

Para responder a esa pregunta hicimos un estudio preliminar sobre parte de los alumnos de 2009 en donde analizamos las diferencias en el IC entre los participantes masculinos y femeninos. Pudimos estudiar solamente un número reducido de alumnos, 88 en total, con 57 mujeres y 31 varones. Los resultados, que exponemos en la tabla 4.141, mostraron que no había diferencias significativas entre mujeres y hombres en cuanto a la tendencia a asumir riesgos en el examen.

TABLA 4.141.
IC en varones y mujeres (2009)

	<i>varones</i>	<i>mujeres</i>
Número de sujetos analizados	31	57
Nota media en el test (según aciertos)	6.8	7.7
I.C. medio	1.06	1.02
Desviación típica	0.21	0.18
Significación estadística	F = 0.8806, t = 0.896 p = 0.187 (ns)	

Como el número de participantes no fue muy elevado, decidimos volver a repetir este estudio en el presente trabajo. Para ello dispusimos de los estudiantes del año 2014, de Enfermería y Fisioterapia de segundo curso. En total fueron 139 estudiantes, con un fuerte predominio femenino: 109 mujeres frente a 31 hombres. Analizamos no solamente el IC sino el PC-aciertos, medida más ajustada de la tendencia al riesgo. Los datos obtenidos los exponemos en la tabla 4.142, donde se muestra la correlación del

género de los estudiantes y sus puntuaciones en confianza e índices relacionados –IC, PC-aciertos-.

TABLA 4.142.

Relación de la confianza con el género de los estudiantes.

	PC	IC	PC-aciertos
Género masculino	-0.148	-0.015	-0.041
Significación estadística	NS	NS	NS

Nota. *r* de Pearson. Estudio realizado en 139 estudiantes. Nivel de significación de la *r* de Pearson para una población de 139 sujetos: *r* de 0.167, $p < 0.05$; *r* de 0.219, $p < 0.01$; *r* de 0.281, $p < 0.001$; *r* de 0.357, $p < 0.0001$.

Estos datos indican que no hay correlación entre los resultados en PC, IC y PC-aciertos relacionada con el género. En la tabla 4.143 completamos estos datos añadiendo las medias de las puntuaciones de los alumnos divididas en dos cohortes: alumnos del género femenino y masculino. Este estudio de cohortes nos permite calcular las medias en los parámetros estudiados en cada sexo; confirma los datos encontrados en el estudio de correlaciones. El IC es ligeramente más elevado y la PC-aciertos ligeramente menor en las estudiantes mujeres que en los estudiantes varones, pero con diferencia mínima y sin ninguna significación estadística.

TABLA 4.143.

Datos estadísticos de la PC, IC y PC-aciertos.

sexo	N	PC	δ	Error T.	d	F	p	t	p
mujer	108	6.23	1.553	0.150	-0.36	0.244	>0.05	1.714	>0.05
varón	31	5.64	1.922	0.345			NS		NS
sexo	N	IC	δ	Error T.	d	F	p	t	p
mujer	108	1.08	0.575	0.056	-0.03	0.004	>0.05	0.165	>0.05
varón	31	1.06	0.574	0.103			NS		NS
sexo	N	PC-aciertos	δ	Error T.	d	F	p	t	p
mujer	108	-0.22	1.426	0.138	0.10	0.047	>0.05	0.485	>0.05
varón	31	-0.08	1.482	0.266			NS		NS

Nota. Estudio realizado en los alumnos de 2º de Enfermería y Fisioterapia en febrero-marzo de 2014.

Como se ve, el índice de confianza, que mide la tendencia al riesgo, es ligeramente más alto en las estudiantes femeninas que en los masculinos (1.08 frente a 1.06), pero la diferencia es ínfima y no significativa. Igualmente, el coeficiente PC-aciertos es muy parecido en ambos géneros, muy próximo a 0 -0.22 en mujeres y -0.08 en varones. No encontramos datos que nos digan que uno de los sexos es significativamente más arriesgado, o tenga más autoconfianza, en este tipo de examen.

Si ambos géneros adecúan de manera semejante su nivel de confianza a sus conocimientos, es lógico que al evaluar la habilidad del estudiante con el TBC, medida según la diferencia [Nota TBC-Nota aciertos], no haya tampoco diferencias entre los sexos, como se ve en la tabla 4.144.

TABLA 4.144.

TBC-aciertos, comparación mujeres-hombres.

sexo	N	TBC-aciertos	δ	Error T.	d	F	p	t	p
mujer	108	-0.01	0.622	0.060	-0.08	0.091	>0.05	0.506	>0.05
varón	31	-0.06	0.617	0.111			NS		NS

Nota. Estudio realizado en los alumnos de 2° de Enfermería y Fisioterapia en febrero-marzo de 2014.

En esta tabla se comprueba lo que ya vimos en el estudio de correlaciones, ambos géneros manejan el TBC con similar habilidad, ajustando adecuadamente, por término medio, su elección del grado de confianza a sus conocimientos. No hay diferencias estadísticamente significativas entre los estudiantes varones y mujeres en nuestro grupo.

En nuestro estudio, el TBC no presenta sesgo de género, ni en cuanto a la habilidad del alumno con este sistema de examen ni en cuanto a una diferente asunción de riesgos por cada género. Nosotros no encontramos que los estudiantes sean más o menos arriesgados que las estudiantes.

5. CONCLUSIONES

Tal como señalamos al exponer los objetivos del presente estudio, un buen sistema de evaluación, para ser un instrumento útil al profesor y al estudiante, ha de cumplir varias condiciones. Enumeramos las siguientes:

- **Validez y fiabilidad.** Ha de proporcionar una medida fiel y discriminativa de los conocimientos del alumno.
- **Autoevaluación.** Ha de proporcionar al alumno información válida y comprensible sobre su aprendizaje y ayudarle a orientar el proceso.
- **Aceptación por el alumno.** No ha de generar rechazo o un excesivo estrés en el alumno, al menos en comparación con otros tipos de evaluación.
- **Reflexión.** Debe inducir en el estudiante un proceso de reflexión sobre sus conocimientos y su aprendizaje.
- **Motivación.** Debe estimular el esfuerzo de aprendizaje del alumno, no dándole expectativas de excesiva facilidad para superar la materia.
- **Aprendizaje profundo.** Fomentará las estrategias de aprendizaje profundo, frente a las superficiales; o, dicho de otra manera, debe distinguir entre los aprendices profundos y los superficiales de tal modo que los primeros resulten beneficiados y los segundos penalizados (Tian, 2007)
- **Ausencia de sesgos.** No ha de suponer una mayor dificultad/facilidad a ningún grupo de alumnos en función de circunstancias extra académicas; específicamente, no ha de tener sesgo de género en sus resultados.

En el caso concreto de los sistemas de evaluación basados en el grado de confianza, Gardner–Medwin (2003) añade algunas exigencias más:

- **Simplicidad.** El alumno ha de entender rápidamente las bases del sistema de evaluación y la puntuación en función de la confianza.
- **Motivación a la sinceridad.** El alumno ha de entender que la sinceridad a la hora de escoger su opción de seguridad es premiada con una puntuación más elevada.
- **Flexibilidad.** Que se pueda aplicar a cualquier formato en que se le planteen al alumno respuestas correctas e incorrectas.
- **Coherencia.** Cada opción de confianza que se le propone al alumno ha de corresponder a una situación en la que su elección sea la más racional.

Todas estos requisitos han sido verificadas por nosotros, lo que confirma la excelencia de este sistema de examen, y, además, hemos encontrado que constituye un útil instrumento para estudiar características del estudiante, como su tendencia a la prudencia o a asumir riesgos, que pueden ser útiles en el proceso de orientación del aprendizaje y de preparación para un futuro profesional.

A continuación exponemos cómo el TBC cumple esas condiciones, según los datos que hemos expuesto en el capítulo «Resultados».

5.1. VALIDEZ DEL TBC

Los resultados expuestos en el capítulo 4.1.1. demuestran que el TBC tiene una validez superior al examen tradicional corregido y puntuado teniendo en cuenta solo el número de aciertos. Lo comprobamos correlacionando la nota en el test, corregida por dos procedimientos –basado en la confianza y por aciertos- con la de un examen simultáneo y sobre el mismo temario del tipo de preguntas abiertas. Cuando el test lo

corregimos según el sistema TBC la correlación fue de 0.705, mientras que corregido teniendo en cuenta solo el número de aciertos la correlación fue de 0.672; aunque la diferencia fue pequeña fue estadísticamente significativa (t de Student, $p < 0.05$). Así pues, **al utilizar el sistema de test basado en la confianza aumentamos la validez de los resultados del examen.**

5.2. CONSISTENCIA INTERNA DEL TBC

Hemos comparado la consistencia interna de nuestros exámenes corrigiéndolos de dos maneras, según la manera tradicional, que solo tiene en cuenta el número de aciertos, y según el nivel de confianza, encontrando que el TBC proporciona una mayor fiabilidad que la puntuación por aciertos: en cinco de los seis casos estudiados la consistencia interna era mayor con el TBC y en el sexto ambas eran iguales. Si bien se puede objetar que al aumentar el número y rango de puntuaciones en cada pregunta con el TBC frente al sistema tradicional se facilitarían que el índice alfa de Cronbach fuese superior, por el aumento de las varianzas, por otro lado, al evaluar el TBC dos factores (conocimientos y confianza) en lugar de uno solo (conocimientos), se tendería a reducir la consistencia interna. Nosotros consideramos, al plantear el presente trabajo, que sería suficiente demostrar que la consistencia interna no se reduce al utilizar el TBC.

Los índices alfa de Cronbach de los seis grupos estudiados eran, de media, de 0.768 con el TBC y de 0.662 con el test convencional *por aciertos* (resultados expuestos en la tabla 4.5). **Aplicar la medida de la confianza a la simple medida del conocimiento mejora, en nuestra experiencia, la fiabilidad del examen.**

5.3. ACEPTACIÓN DEL TBC

El examen de tipo test con puntuación basada en el grado de seguridad (TBC) en cuanto a las preferencias de los alumnos merece un «aprobado»; en una escala de 1 a 6 recibe de puntuación un 3.66. Es un examen que gusta menos a los alumnos que el test tradicional, con puntuación basada en el número de aciertos (que obtiene un 4.88), pero más que el examen de preguntas abiertas (que es valorado con un 2.49). Los alumnos aceptan el TBC; les gusta más que el examen de preguntas abiertas pero menos que el test por aciertos. Podemos ver los resultados comparados de la aceptación de los tres tipos de examen en la tabla 4.6 (que expusimos en el capítulo de Resultados).

Tabla 4.6.

Ítem 2, 3 y 4: preferencia de los alumnos por cada examen.

	Test convencional	TBC	Preguntas abiertas
MEDIA (de 1 a 6)	4.88	3.66	2.49
Porcentaje sobre la puntuación máxima	77.6 %	53.2 %	19.8%
	F = 174,39	p<0.001	Eta ² = 0.312

El porcentaje de alumnos que expresaron que les gustaba el TBC (en los tres niveles de preferencia positiva) fue de un 56.6 %, frente al 43.4 % que expresaron una opinión negativa (Tabla 4.5). El TBC no provoca un rechazo en la mayoría de los alumnos, aunque su nivel de aceptación es inferior al del test convencional por aciertos al que los estudiantes están acostumbrados. No hay rechazo generalizado de este último examen, las valoraciones positivas y las negativas están equilibradas y el TBC es preferido, en cualquier caso, al examen de preguntas abiertas, que es, por término medio, el menos apreciado con diferencia notable. En esta pregunta no parece que los estudiantes respondan buscando la *conformidad* con el profesor, pues no dudan en manifestar su preferencia clara por el test por aciertos –la encuesta es realizada por el

profesor que utiliza el TBC y cabría, en consecuencia, temer que los resultados fuesen afectados por el *sesgo de la conformidad* a favor de su examen, pero en esta pregunta ese sesgo no es evidente-; en las siguientes preguntas, que comparan el TBC con los otros dos exámenes, aunque los participantes hayan manifestado en su mayoría su preferencia por el test por aciertos, otorgan más puntuación al TBC, siendo este un dato que apoya la sensación que las respuestas son sinceras.

Los alumnos que tienden a otorgar una mayor puntuación en preferencia al test basado en la confianza son aquellos con mayor nivel de conocimientos, con mejores expectativas de resultados académicos, con enfoques profundos de aprendizaje-sobre todo estrategias profundas- con mayor autorregulación y que consideran que han sido hábiles utilizando este sistema de examen. Por el contrario, los alumnos con enfoques superficiales de aprendizaje tienden a otorgar una puntuación más baja, en cuanto a preferencias, al TBC (datos expuestos en la tabla 4.6). Podemos afirmar que los alumnos con mejores hábitos de estudio son los que muestran mayor preferencia por el test basado en la confianza.

Un factor relacionado con la aceptación de un sistema de examen es el estrés que genera en los alumnos. Una de las preguntas, la correspondiente a los ítems 26, 27 y 28, incidía en este punto: «¿Te sientes muy estresado antes de hacer un examen?»; la puntuación media según cada tipo de test fue de 4.31 en el test por aciertos, de 4.81 en el TBC y de 4.92 con el examen de preguntas abiertas (en una escala de 1 a 6). Al igual que sucedía con la preferencia, los alumnos reconocen padecer menor estrés cuando se enfrentan a un test convencional que cuando se enfrentan a un test basado en la confianza, siendo el examen de preguntas abiertas el que genera más estrés. Las diferencias en este punto entre los diferentes tipos de exámenes no son muy marcadas, como vimos en la tabla 4.64.

TABLA 4.64.

Estadísticos básicos del *estrés* con cada tipo de examen (ítems 26, 27 y 28)

	Test por aciertos	TBC	Preguntas abiertas
MEDIA (de 1 a 6)	4.31	4.81	4.92
Desviación estándar	1.31	1.18	1.29
Porcentaje sobre puntuación máxima	66.2 %	76.2 %	78.4 %
Significación de la diferencia con el TBC	p <0.0001	-	p <0.05
F = 56.167	P<0.001	Eta ² = 0.172	

5.4. MODIFICACIÓN EN LOS HÁBITOS DE ESTUDIO

Los estudiantes afirman que el tipo de examen que esperan condiciona su manera de estudiar de forma importante; la puntuación media en esta pregunta era de 4.6 en una escala de 1 a 6 (tabla 4.10). El 76.8 % opinan que el tipo de examen les condiciona el estudio bastante, mucho o totalmente; frente a ellos, solo el 13.6 % opinan que les condiciona muy poco o nada.

¿En qué sentido modifica el TBC la forma de estudiar, frente al test convencional por aciertos y frente al examen de preguntas abiertas?

Los resultados concernientes a estas cuestiones los expusimos en las tablas 4.11, 4.16 y 4.20. En ellas vimos que los alumnos opinaban que el TBC obligaba a aumentar el tiempo dedicado al estudio frente al test por aciertos (tabla 4.11), exigía mayor esfuerzo de comprensión en su preparación que los otros dos tipos de examen (tabla 4.16) y también obligaba a realizar mayor esfuerzo de memorización que el test por aciertos (tabla 4.20). En comparación con el examen de preguntas abiertas, el TBC permitía dedicar menos tiempo al estudio y menor esfuerzo a la memorización de datos,

pero exigía un mayor esfuerzo de comprensión (tablas 4.11, 4.12, 4.16, 4.17, 4.20 y 4.21).

Por lo tanto, en opinión de los estudiantes, cuando añadimos el grado de seguridad a un examen de tipo test, conseguimos que dediquen más tiempo al estudio y que dediquen un mayor esfuerzo a comprender lo que memorizan.

Es opinión común de los estudiantes que en un examen de tipo test, para obtener un buen resultado, no hace falta tener un dominio de la materia tan completo como cuando el examen es de preguntas abiertas; bastaría tener una idea general de la materia, pues es menos exigente reconocer la respuesta correcta, redactada ya entre varias falsas, que tener que redactar dicha respuesta partiendo de 0. Nuestra encuesta confirma esta impresión en la opinión que exponen los estudiantes; se observa en sus respuestas sobre el tipo de estudio que genera cada tipo de examen, en la comparación entre el test *por aciertos* y el examen de *preguntas abiertas*: ellos opinan que no hace falta dedicar tanto tiempo al estudio y a la memorización cuando el examen es del primer tipo mientras que no ocurre así cuando es del segundo. Al añadir la cuestión de la *confianza*, con el TBC, conseguimos que los estudiantes aumenten su esfuerzo en la preparación del examen. Ese esfuerzo se debe a un mayor tiempo dedicado al estudio, a un mayor esfuerzo por entender lo que se estudia y a un mayor esfuerzo de memorización frente al *test por aciertos*, siendo destacada la opinión de que el TBC estimula el esfuerzo de comprensión más que los otros dos exámenes.

La cuestión del «esfuerzo de comprensión» entendemos que es la más relevante en la preparación de un examen; dicha preparación, especialmente en una materia predominantemente de conocimientos, tal como son las que el autor imparte, se puede

dividir en tres subtareas: reunir la información, comprender los conceptos relevantes y memorizarlos. La memorización está actualmente desprestigiada, pero es necesaria, pues la mera comprensión no garantiza la conservación de los conocimientos y menos aún su reproducción. La memoria es una facultad plástica y adaptable, y su ejercicio mejora en todos los sentidos las capacidades cognitivas; no la debemos desdeñar. Aceptado esto, la tarea a la que damos más importancia en el proceso de aprendizaje es la de **la comprensión** de los conceptos. La comprensión profunda es lo que da validez al aprendizaje, lo que lo hace útil. Nuestros alumnos asocian claramente la comprensión con el buen aprendizaje, como se ve en la tabla 4.18, donde -en cada tipo de examen- los sujetos que opinan que ese examen «*obliga a comprender mejor*» son los que le dan mayor puntuación en la pregunta «*me hace aprender más*» (correlaciones de 0.51, 0.37 y 0.56, respectivamente para los tres tipos). En ese sentido, el que los alumnos opinen que tener que responder a la cuestión de la seguridad les obliga a tener que comprender mejor todo, en comparación con los otros tipos de examen, hace destacar la eficacia del TBC como método docente.

En la opinión expresada por los estudiantes, el TBC les hace prepararse mejor para los exámenes que los test convencionales por aciertos. Esta opinión se da en mayor grado en los alumnos con motivos de aprendizaje profundos (r de 0.32) y con mejor autorregulación (r de 0.26), y es menos frecuente en los alumnos con motivos superficiales (r de -0.23).

5.5. OPINIÓN DE LOS ALUMNOS SOBRE LAS CUALIDADES DEL TBC EN COMPARACIÓN CON LOS DOS OTROS TIPOS DE EXAMEN

Como dijimos antes, es condición necesaria para un buen sistema de evaluación que sea bien aceptado por el alumno; esa aceptación no solo ha de ir en el sentido de los gustos del alumno, sino en su opinión sobre la validez del examen como sistema de evaluación. Con independencia de sus preferencias personales, si un sistema de evaluación es reconocido como bueno por un alumno será mejor aceptado por él.

Para evaluar la opinión de los alumnos sobre las virtudes de cada examen, su «*validez autopercebida*», en lugar de plantear una pregunta directa como «*¿qué tipo de examen crees que es mejor?*», que juzgamos estaría influida por el gusto personal expresado en los tres primeros ítems, preferimos subdividir esa cuestión en tres preguntas: *¿qué tipo de examen le permite al profesor conocer mejor lo que sabes?*, *¿qué tipo de examen te permite darte cuenta mejor de tus lagunas formativas?* y *¿qué tipo de examen te hace aprender más?*. En esas preguntas condensábamos la opinión del alumno sobre la *validez*, tal como él la percibe, del sistema de examen, sobre su utilidad para la *autoevaluación* y sobre el impacto, la *potenciación*, que generaba en su aprendizaje. Esas preguntas se analizaban individualmente y también con ellas construimos un agrupamiento que valoraba lo que hemos llamado «*validez percibida por el estudiante*» con cada tipo de examen.

En las tres cuestiones, *validez*, *autoevaluación* y *potenciación del aprendizaje*, el TBC recibía una puntuación más alta que los otros dos sistemas testados. Recordemos los valores expresados en las tablas 4.22, 4.23 y 4.24:

TABLA 5.1.

Media de los ítems que definen el constructo «*validez auto percibida*».

Ítems que definen el constructo: <i>validez autopercebida</i>	Test <i>por</i> <i>aciertos</i>	TBC	Preguntas <i>abiertas</i>	F de Ledene	p
15-16-17: Permite al profesor conocer mejor lo que sé	3.93	5.03	4.67	54.99	<0.001
18-19-20: Me hace descubrir lagunas formativas	4.07	4.98	4.45	42.27	<0.001
21-22-23: Me hace aprender más	4.39	5.07	4.48	29.72	<0.001
MEDIA de las tres (de 1 a 6)	4.13	5.03	4.53	77.99	<0.001
Media en porcentaje sobre la nota máxima (1-6)	62.6 %	80.5 %	70.6 %		
Significación de la diferencia con el TBC	p <0.001	-	p <0.001		

Nota. Extraída de los estadísticos básicos de esos ítems expuestos en las tablas 4.22, 4.23 y 4.24.

Por lo tanto, los alumnos consideran que el TBC es el mejor sistema de examen de los tres analizados. Se destaca el gran incremento en la opinión del alumno sobre la validez del examen cuando al sistema de test se le añade la valoración de la seguridad: pasa de una puntuación media de 4.13 (62.6 % de la máxima) a una de 5.03 (80.5 %), un tamaño del efecto de un 28.6 %. Esta buena opinión es más relevante si tenemos en cuenta que los alumnos prefieren el test *por aciertos*, les gusta claramente más (asignan en cuanto al gusto una puntuación de 4.88 al test *por aciertos* frente al 3.66 que otorgan al TBC); reconocer que el TBC es mejor, aunque les guste menos, da solidez a esa opinión y elimina la sospecha de *conformidad* con el profesor en la respuesta. Un detalle que refuerza esta opinión positiva sobre el TBC, con independencia del gusto del alumno, es el estudio de la correlación que hay entre creer haber estado acertado con el TBC (Ítem 24) y creer que el TBC es un sistema válido de examen; hay naturalmente una correlación significativa, r de 0.211 (tabla 4.40), aunque menor que la que tiene la relación entre la predilección personal por el TBC y la creencia de haber estado acertado con él: r de 0.534 (tabla 4.41). En consecuencia, el TBC es considerado por los

estudiantes de nuestra encuesta el mejor sistema de examen, aunque no sea el que más les guste.

La idea de que un buen sistema de examen, además de ser válido como instrumento de medida de conocimientos, obliga al alumno a trabajar más preparándose la materia no solo es una opinión propia del docente, que naturalmente desea que el alumno medio trabaje más en su materia, sino que aparece reflejada en nuestra encuesta, pues hay una clara correlación entre la opinión favorable sobre un tipo de examen y el reconocer que se dedica mucho tiempo a prepararlo. Como vemos en la Tabla 5.2 (resumida de la tabla 4.35 de resultados), los alumnos asocian estímulo al esfuerzo con opinión positiva sobre la validez del examen. Por lo tanto, los estudiantes tienden a considerar que obligar a estudiar más es un atributo de un buen sistema de evaluación.

TABLA 5.2.

Correlaciones con los ítems que valoran el tiempo dedicado al estudio con cada examen con la *validez autopercebida* de cada uno de ellos.

	TBC	<i>Test por aciertos</i>	<i>Preguntas a desarrollar</i>
Correlación entre la opinión positiva sobre la <i>validez autopercebida</i> del examen y el hecho de <i>tener que estudiar más con él</i> .	0.382 (p<0.0001)	0.382 (p<0.0001)	0.266 (p<0.001)

Nota. Tabla resumida de la tabla 4.35. r de Pearson.

Otro parámetro que hemos visto se asocia fuertemente a la opinión sobre la bondad de un tipo de examen es la opinión sobre si dicho tipo de examen obliga al alumno a comprender bien lo que se estudia (ítems 8, 9 y 10). Como vimos en el capítulo de «Resultados», hay una asociación muy fuerte entre estos conceptos; lo vemos en la tabla 5.3, que resume la tabla 4.35 de «Resultados».

TABLA 5.3.

Correlación de la validez de un examen con la necesidad de comprender bien los conceptos estudiados.

	TBC	Test por aciertos	Preguntas a desarrollar
Correlación entre la opinión positiva sobre cada tipo de examen y tener que comprender mejor con él.	0.415 (p<0.00001)	0.550 (p<0.00001)	0.647 (p<0.00001)

Nota. Tabla resumida de la tabla 4.35. r de Pearson.

Dadas las altas correlaciones que hemos encontrado entre los tres parámetros que hemos considerado definitorios de la *validez autopercebida* de un examen (lo que sería la medida de la validez del examen tal como es percibida por el estudiante) -*validez de la información que proporciona al profesor, autoevaluación y potenciación del aprendizaje*- y los otros dos que acabamos de exponer - «¿Qué examen te obliga a dedicar más tiempo al estudio?» (ítems 5, 6 y 7), y «¿Qué examen te obliga a entender mejor lo que aprendes?» (ítems 8, 9 y 10)- formamos, para cada tipo de examen, un nuevo agrupamiento quíntuple, que mediría tanto la validez objetiva del examen como las modificaciones positivas que generaría en las actividades de estudio del alumno. A este agrupamiento de 5 preguntas, que valora tanto la validez del examen como la influencia positiva que el examen tiene en el estudio del alumno, lo hemos llamado «*eficacia docente*», pues analiza no solo lo válido que un examen es como sistema de monitorización del aprendizaje, sino también su papel como estímulo al aprendizaje. El estudio factorial exploratorio nos confirmó la validez de este agrupamiento para los tres tipos de examen, aunque en uno de los tipos, en el test por aciertos, a estas cinco preguntas se añadía una sexta, referente al estímulo a la memorización que el examen suponía -algo que no sucedió con los otros tipos-. Por ello nos hemos quedado con este constructo, esta escala de 5 preguntas, como un instrumento útil para valorar lo que

entendemos que son las cualidades positivas de una evaluación desde el punto de vista del alumno.

Las 5 preguntas que forman este agrupamiento, la miniescala de *eficacia docente*, son:

- 1- ¿En qué medida cada tipo de examen te obliga a dedicar más tiempo al estudio?**
- 2- ¿En qué medida cada tipo de examen te obliga a entender bien todo lo que estás estudiando?**
- 3- ¿Qué tipo de examen crees que le permite al profesor conocer mejor lo que el alumno sabe?**
- 4- ¿Qué tipo de examen te hace darte cuenta mejor de tus lagunas formativas?**
- 5- ¿Qué tipo de examen crees que te hace aprender más?**

Este factor, «eficacia docente del examen», tiene una aceptable consistencia interna, con una α de Cronbach de 0.78, 0.74 y 0.89 para cada uno de los tres tipos de examen analizados –test por aciertos, TBC y examen de preguntas abiertas-. Hay que tener en cuenta que estos valores de consistencia interna son menores que los reales –en los dos tipos de test, por aciertos y TBC- debido al formato que hemos escogido en nuestra encuesta, en donde tras cada pregunta se le hacía al alumno puntuar los tres tipos de examen, estableciendo una especie de competencia entre ellos –cuál era mejor, cuál era peor-, lo que hacía que la puntuación que se le daba a uno condicionara la que se le daba a los otros dos, según las preferencias del estudiante; en la tabla número 4.55 expusimos la correlación entre las opiniones sobre la eficacia docente con los tres tipos de examen, encontrando que había una importante correlación entre los dos test (TBC y *test por aciertos*) pero que no había correlación significativa entre la opinión sobre

eficacia docente de los test con el examen de *preguntas abiertas*, tal como exponemos a continuación:

- Correlación entre los valores de *eficacia docente* del TBC y del *test por aciertos*: **0.361** ($p < 0.0001$)
- Correlación entre los valores de *eficacia docente* del TBC y del examen de *preguntas abiertas*: 0.114 ($p > 0.05$, no significativo)
- Correlación entre los valores de *eficacia docente* del *test por aciertos* y del examen de *preguntas abiertas*: 0.021 ($p > 0.05$, no significativo)

Hay, por lo tanto, una correlación importante entre las puntuaciones otorgadas por cada alumno en el factor «eficacia docente» al *test por aciertos* y al TBC; por el contrario, no hay correlación significativa entre los exámenes de tipo test y el de *preguntas abiertas*; el examen de *preguntas abiertas* se muestra casi independiente, en lo que al factor «eficacia docente» se refiere, de los otros. Esta correlación entre el *test por aciertos* y el TBC tiende a disminuir la consistencia interna de la escala de estos dos exámenes, al introducir una variable externa añadida a la de la pregunta básica de la opinión sobre el examen –la variable nueva es la comparación con el otro examen- y esa nueva variable externa distorsiona el cálculo de la α de Cronbach, tendiendo a disminuir sus valores. Por ese motivo, en nuestra opinión, la consistencia interna de la escala «eficacia docente» obtenida con el examen de preguntas abiertas es notablemente mayor, con un α de 0.89 –una cifra elevada-, al no estar condicionada la respuesta a la de los otros dos exámenes. Eso nos indica que si, en lugar de haber comparado los tres exámenes en cada pregunta, hubiésemos hecho la encuesta aisladamente con cada tipo de examen, la consistencia interna del constructo «eficacia docente» con el TBC y el *test por aciertos* hubiera sido más elevada, probablemente mayor de 0.80, tal como se

obtuvo en la escala del examen de preguntas abiertas -0.891-. No era nuestra intención analizar aisladamente cada tipo de examen, sino compararlos; por eso escogimos el formato de preguntas que adoptamos, aunque fuera en detrimento de la fiabilidad de este constructo –que, de todas maneras, ha sido bastante alta para ser factores de solo 5 elementos-.

De esta manera hemos elaborado un constructo que sirve para medir las características positivas de cualquier tipo de evaluación desde el punto de vista del estudiante. Aunque nuestra intención, al diseñar el estudio, no era la de elaborar una escala de validez general, sino solo una encuesta *ad hoc* para analizar la opinión de los alumnos sobre el TBC, hemos obtenido con estas cinco preguntas un constructo que puede ser útil para estudiar cualquier tipo de prueba evaluatoria. Sin duda habría que refinar algo más el instrumento, aumentar su fiabilidad y, posiblemente, añadir alguna pregunta más, lo que sobrepasa los objetivos del presente trabajo, pero tal como está creemos que es una escala útil para medir las cualidades positivas de un examen, y puede ser utilizado en otros estudios. Las cinco preguntas del factor «eficacia docente», tal como hemos expuesto en «Metodología» son:

- **¿En qué medida cada tipo de examen te obliga a dedicar más tiempo al estudio?**
- **¿En qué medida cada tipo de examen te obliga a entender bien todo lo que estás estudiando?**
- **¿Qué tipo de examen crees que le permite al profesor conocer mejor lo que el alumno sabe?**
- **¿Qué tipo de examen te hace darte cuenta mejor de tus lagunas formativas?**

- **¿Qué tipo de examen crees que te hace aprender más?**

Estas preguntas están formuladas según el formato de nuestra encuesta, en la que comparamos constantemente los exámenes de tipo *test por aciertos*, TBC y *preguntas abiertas*, pero se pueden adaptar, modificando ligeramente la redacción, a la evaluación –no comparativa- de cualquier tipo de examen o prueba evaluatoria. En tal sentido, podrían quedar redactadas de una manera similar a la siguiente:

- **¿En qué medida *el examen/evaluación ...* te obliga a dedicar mucho tiempo al estudio?**
- **¿En qué medida *el examen/evaluación ...* te obliga a entender bien todo lo que estás estudiando?**
- **¿En qué medida crees que *el examen/evaluación ...* le permite al profesor conocer lo que el alumno sabe?**
- **¿En qué medida *el examen/evaluación ...* te hace darte cuenta mejor de tus lagunas formativas?**
- **¿Crees que *el examen/evaluación ...* te hace aprender más?**

El nuevo constructo lo hemos denominado «eficacia docente», pues no solamente evalúa la *validez percibida* por el alumno de un modelo de examen, sino que evalúa también la opinión del alumno sobre la *modificación que induce en sus hábitos de estudio*, y sobre el progreso que ese tipo de examen induce en su aprendizaje. Por lo tanto, este constructo mide la «calidad» o excelencia de un sistema de evaluación. Creemos que podría ser útil en futuros estudios de la opinión de los estudiantes sobre los exámenes y sistemas de evaluación.

Estudiando los resultados obtenidos con estos agrupamientos -*validez autopercebida* y *eficacia docente* del examen- los resultados anteriores se confirman y potencian: El TBC es el sistema que obtiene mayor puntuación de los tres. Por lo tanto, es el examen mejor conceptualizado por los alumnos por su validez y, globalmente, el que más modifica en sentido positivo su estudio y aprendizaje (solo en el apartado concreto de «tiempo dedicado al estudio» el examen de *preguntas abiertas* tiene puntuación superior). En definitiva, los estudiantes encuestados opinan que el TBC es el examen con mayor *eficacia docente*; dicho de otra manera: el que más ajustadamente valora lo que saben y el que más les estimula a un aprendizaje de calidad. La puntuación que otorgan en este factor se muestra en la tabla 5.4.

TABLA 5.4.

«Eficacia docente» de cada tipo de examen (modificada de la 4.58).

TIPO DE EXAMEN	Puntuación media (de 1 a 6)	Porcentaje sobre puntuación máxima
<i>Eficacia docente</i> , TBC (6+9+16+19+22)	5.09	81.8 %
<i>Eficacia docente</i> , test por número de aciertos(5+8+15+18+21)	4.28	65.6 %
Eficacia docente, examen de preguntas abiertas(7+10+17+20+23)	4.73	74.6 %
F = 77.99 p < 0.001 Eta ² = 0.225		

Nota. Todas las diferencias son significativas con un $p < 0.001$. Tabla resumida de la 4.58.

En comparación con el test *por aciertos*, el TBC obtiene una puntuación un 24.7 % superior (81.8 % frente a 65.6 %). Los estudiantes han opinado en la encuesta que cuando introducimos la pregunta sobre el grado de seguridad, la validez del examen mejora en todos los sentidos; les obliga a estudiar más y a realizar un mayor esfuerzo de comprensión de lo que estudian.

Esta opinión favorable al TBC como buen sistema de examen se da más en los alumnos con enfoques profundos de aprendizaje (r de 0.35), especialmente *motivos*

profundos (r de 0.36), y en los que tienen mayor autorregulación (r de 0.23); y es menos frecuente en aquellos que tienen enfoques superficiales de aprendizaje (r de -0.23), especialmente motivos superficiales (r de -0.27), tal como se ve en la Tabla 4.52. Tener una buena opinión del TBC aparece asociado a enfoques profundos de aprendizaje, mientras que una opinión menos favorable a este sistema de examen se asocia a los superficiales. En las tablas de correlaciones entre la opinión sobre un tipo de examen y los enfoques de aprendizaje resulta llamativo que los exámenes de tipo test por aciertos y de preguntas abiertas no muestran correlaciones significativas. Solo el TBC se correlaciona con ellos: tener enfoques profundos con una opinión positiva sobre ese examen y tener enfoques superficiales con una opinión negativa. Volvemos a poner la Tabla 4.58 para apreciar esto con toda su evidencia:

TABLA 4.58.

Correlación del factor «eficacia docente» con enfoques de aprendizaje y autorregulación en los tres tipos de examen.

	<i>Eficacia docente TBC</i> 6+9+16+19+22	<i>Eficacia docente test por aciertos</i> 5+8+15+17+21	<i>Eficacia docente preguntas a desarrollar</i> 7+10+17+20+23
Ítems de ENFOQUE PROFUNDO	0.354 (p<0.0001)	NS	0.14 (p<0.05)
Ítems de MOTIVO PROFUNDO	0.360 (p<0.0001)	NS	0.15 (p<0.05)
Ítems de ESTRATEGIA PROFUNDA	0.278 (p<0.0001)	NS	NS
Ítems de ENFOQUE SUPERFICIAL	-0.228 (p<0.001)	NS	NS
Ítems de MOTIVO SUPERFICIAL	-0.273 (p<0.0001)	NS	NS
Ítems de ESTRATEGIA SUPERFICIAL	-0.15 (p<0.05)	-0.13 (p<0.05)	NS
Ítems de AUTORREGULACIÓN	0.233 (p<0.001)	0.12 (p<0.05)	0.13 (p<0.05)

Nota. Se resaltan las correlaciones mayores de 0.202, que tienen una p<0.001 (r de Pearson).

En resumen, los alumnos opinan que el TBC es el mejor sistema de evaluación de los tres encuestados en todos los sentidos que definen las cualidades positivas que ha de tener un examen –a las que hemos denominado «eficacia docente»-. Es decir, da mejor información al profesor sobre los conocimientos del alumno y a este sobre su propio aprendizaje; hace que el estudiante dedique más tiempo al estudio, que se esfuerce en comprender más lo que estudia, y, por último, es el sistema de examen que le hace aprender más, en opinión del propio alumno.

5.6. ESTUDIO DE LOS ENFOQUES DE APRENDIZAJE DE LOS ALUMNOS Y SU CORRELACIÓN CON LA OPINIÓN SOBRE LOS DIVERSOS TIPOS DE EXAMEN Y CON SUS RESULTADOS CON EL TBC

Nosotros hemos estudiado la relación entre los enfoques de aprendizaje, tanto profundos como superficiales, con los resultados académicos en nuestras asignaturas así como con la habilidad del estudiante con el TBC y las opiniones de este sobre los diversos tipos de exámenes.

Para hacer el estudio de los enfoques de aprendizaje, añadimos a nuestra encuesta la escala de Biggs R-SPQ-2, de 20 ítems que mide dos factores, los enfoques profundo y superficial. Obtuvimos unas excelentes medidas de consistencia interna con el R-SPQ-2F en nuestra encuesta, una α de Cronbach de 0.804 para los enfoques de aprendizaje profundo y de 0.760 para los superficiales, lo que confirma la adecuación de este test como instrumento de medida.

Un dato que apareció cuando hicimos el análisis factorial de nuestra escala R-SPQ-2F fue el de que, cuando el análisis se ajustaba a dos componentes, se formaban claramente los agrupamientos de ítems de enfoques profundos y los de enfoques superficiales (tabla 4.118), pero cuando se ajustaba el análisis a un mayor número de componentes, cuatro o cinco, no se separaba el componente «enfoque profundo» en sus dos subfactores «motivo» y «estrategia», sino que se mantenía como un componente único sin distinción entre los ítems de motivación y los de estrategia; y, al mismo tiempo, los enfoques superficiales se dispersaban entre cuatro pequeños subfactores que tampoco se ceñían a la división preestablecida en «motivos» y «estrategias» (tablas 4.119 y 4.120). Entre nuestros estudiantes la escala R-SPQ-2F no se comporta conforme a lo establecido en otros trabajos (Biggs, 2002); aparece clara la división entre enfoques profundos y superficiales, pero la subdivisión de estos dos factores en los subfactores «motivo» y «estrategia» no se confirma. A pesar de ello, dada la aceptación generalizada de estos subfactores de la escala de Biggs, hemos mantenido en nuestro estudio la diferencia entre *estrategias* y *motivos*, la cual ha funcionado notablemente bien cuando se analizaban las correlaciones con otros ítems o factores.

Los alumnos otorgaron mayor puntuación a los enfoques profundos que a los superficiales, como suele serlo habitual con estas encuestas. Las medias de puntuación dadas por los estudiantes a estos factores las mostramos en resultados en las tablas 4.74, 4.75 y 4.76. En resumen:

Enfoques de aprendizaje:

- Profundo, media de **38.63**, $\delta = 7.058$ (3.86 de media por ítem)
- Superficial, media de **25.51**, $\delta = 6.924$ (2.55 de media por ítem)

(F de 299.97, $p < 0.0001$, tamaño del efecto $d = 1.87$)

Motivos de aprendizaje:

- Profundo, media de **20.51**, $\delta = 3.905$ (4.10 de media por ítem)
 - Superficial, media de **11.50**, $\delta = 3.523$ (2.30 de media por ítem)
- (F de 299.97, $p < 0.0001$, tamaño del efecto $d = 2.42$)

Estrategias de aprendizaje:

- Profundo, media de **18.12**, $\delta = 3.928$ (3.62 de media por ítem)
 - Superficial, media de **14.00**, $\delta = 4.262$ (2.80 de media por ítem)
- (F de 299.97, $p < 0.0001$, tamaño del efecto $d = 1.01$)

Todas las diferencias son estadísticamente significativas ($p < 0.0001$). Los alumnos, al evaluar su propio aprendizaje, puntúan más los enfoques profundos que los superficiales. La diferencia entre lo profundo y lo superficial es máxima cuando se comparan motivaciones; de una media de 4.10 por ítem en «motivos profundos» -20.5 en 5 ítems- a 2.30 por ítem en «motivos superficiales» -11.5 en 5 ítems-, un 78% más para motivos profundos (tamaño del efecto de 2.42 veces la DS). En estrategias esta diferencia, notable de todas maneras, es bastante menor; de 3.63 en «estrategias profundas» a 2.80 en «estrategias superficiales», un 30 % más para las estrategia profundas (tamaño del efecto 1.01 veces la DS). Tenemos la impresión de que esta mayor diferencia en los motivos puede obedecer a la tendencia de los alumnos a dar la opinión correcta que se espera de un «buen estudiante», pues los motivos analizan las actitudes frente al aprendizaje, actitudes cuya determinación por el estudiante es subjetiva; por el contrario, las «estrategias» evalúan comportamientos concretos que el estudiante puede reconocer de forma más objetiva.

Analizando las correlaciones de los enfoques de aprendizaje con la encuesta que valoraba la opinión del alumno, constatamos que los alumnos a los que les gusta más el TBC, los que tienen mejor opinión del TBC como sistema de examen, los que modifican más su forma de estudiar en sentido positivo con el TBC y los que se muestran más hábiles con el TBC son los que tienen enfoques de aprendizaje más profundos y dan menos puntuación a los superficiales. Esto se ha indicado antes, en resultados, pero lo resumimos en la tabla 5.6 para los enfoques profundos de aprendizaje.

Tabla 5.6.

Correlaciones de la escala de enfoques profundos con las opiniones sobre el TBC.

	Enfoque profundo	Motivo profundo	Estrategia profunda
Me gusta el TBC (ítem 3)	0.225 (p<0.001)	0.190 (p<0.01)	0.215 (p<0.001)
Me hace estudiar más el TBC (ítem 6)	0.196 (p<0.01)	0.190 (p<0.01)	0.163 (p<0.05)
El TBC me obliga a comprender mejor (ítem 9)	0.303 (p<0.0001)	0.316 (p<0.0001)	0.229 (p<0.001)
<i>Validez autopercibida</i> del TBC (ítems 16+19+22)	0.298 (p<0.001)	0.307 (p<0.0001)	0.229 (p<0.001)
<i>Eficacia docente</i> del TBC (6+9+16+19+22)	0.354 (p<0.0001)	0.360 (p<0.0001)	0.278 (p<0.0001)
Estuve acertado al escoger la opción de seguridad	0.241 (p<0.001)	0.171 (p<0.01)	0.262 (p<0.0001)

Nota. Se destaca en negrita el valor «r» más alto de cada fila, comparando «motivos» con «estrategias». Coeficiente r de Pearson.

En la tabla 5.7 resumimos las correlaciones de los enfoques superficiales de estudio con los ítems que expresan la opinión del alumno sobre el TBC

Tabla 5.7.

Correlaciones de los **enfoques superficiales** con las opiniones sobre el TBC.

	Enfoque superficial	Motivo superficial	Estrategia superficial
Me gusta el TBC (ítem 3)	-0.142 (p<0.05)	NS	-0.168 (p<0.01)
Me hace estudiar más el TBC (ítem 6)	-0.147 (p<0.05)	-0.152 (p<0.05)	NS
El TBC me obliga a comprender mejor (ítem 9)	-0.192 (p<0.01)	-0.231 (p<0.001)	-0.168 (p<0.01)
Validez autopercebida del TBC (ítems 16+19+22)	-0.198 (p<0.01)	-0.245 (p<0.001)	-0.119 (p = 0.05)
Eficacia docente del TBC (6+9+16+19+22)	-0.228 (p<0.001)	-0.273 (p<0.0001)	-0.145 (p<0.05)
Estuve acertado con el TBC	NS	NS	NS

Nota. Se destaca en negrita el valor «r» más alto comparando «motivos» con «estrategias» .r de Pearson.

Vimos en la tabla 5.6 que hay una correlación entre tener enfoques profundos de aprendizaje y tener una *buena opinión* del TBC como sistema de evaluación –gusto por el TBC, opinión de que hace dedicar más tiempo al estudio y entender mejor, *validez autopercebida*, *eficacia docente* y haber estado acertado con él-; esta correlación es mayor en la mayoría de estos ítems con los *motivos profundos* que con las *estrategias profundas* (la excepción, significativa, es con la pregunta que hace referencia a la habilidad del estudiante con el TBC). Creemos que es lógico que haya esta mayor correlación de los ítems que expresan la opinión del alumno sobre el examen con los ítems de *motivo profundo*, que expresan también opiniones del alumno, que con los de *estrategias*, que describen hechos y situaciones más que opiniones. Las opiniones se correlacionan mejor con otras opiniones que con los hechos. Por eso la pregunta sobre lo afortunado que el alumno ha estado con el TBC, en cambio, se correlaciona más con las *estrategias* que con los *motivos*, pues esta pregunta inquiriere sobre hechos de los que el alumno dispone de una información objetiva –los hechos se correlacionan más con otros hechos que con opiniones-.

En la tabla 5.7 vemos que los ítems de enfoques superficiales tienen una correlación negativa con los que expresan opiniones positivas sobre el TBC. Los alumnos que reconocen tener enfoques superficiales tienen peor opinión del TBC que aquellos que tienen enfoques de aprendizaje profundos. También aquí, con los *enfoques superficiales*, se destaca el hecho de que los ítems de opinión del alumno sobre el TBC se correlacionan más con los de *motivos* –opinión- que con los de *estrategias superficiales* –hechos-.

El reconocimiento de los enfoques profundos de aprendizaje es una de las condiciones necesarias para considerar excelente a un sistema de evaluación (Tian, 2007). La habilidad del alumno con el uso del TBC (ítem 24) se correlaciona con los enfoques profundos –más con estrategias- y no se correlaciona, en cambio, con la puntuación a los enfoques superficiales, lo que coincide con lo que encontramos también nosotros cuando estudiamos las correlaciones de los enfoques de estudio con el parámetro *TBC-aciertos*, que mide objetivamente esa habilidad: correlación positiva con enfoques profundos –sobre todo estrategias-, ausencia de correlación con enfoques superficiales. Esta coincidencia de dos series independientes de correlaciones, cada una de ellas con el mismo significado, la correlación del factor *enfoques profundos* con un ítem de una encuesta de opinión en un caso y una puntuación objetiva de un examen –que el alumno realmente no conoce con total exactitud- en el otro, con valores muy próximos en ambos casos, y similar paralelismo en cuanto a los subfactores «*motivos*» y «*estrategias*», da solidez y objetividad a nuestra encuesta: los alumnos con más alta puntuación en *enfoques profundos* –sobre todo *estrategias*- son más hábiles utilizando el TBC que los demás. En la tabla 5.8 mostramos este llamativo paralelismo.

Tabla 5.8.

Correlaciones entre la puntuación en enfoques profundos y los dos ítems que miden la *habilidad* con el TBC.

	Enfoques profundos	Motivos profundos	Estrategias profundas
Ítem 24: «¿Has estado <i>acertado</i> al escoger tu opción de seguridad?»	0.20 (p<0.01)	0.13 (p<0.05)	0.23 (p<0.001)
TBC-aciertos (medida objetiva del «acierto» en la opción de seguridad)	0.24 (p<0.001)	0.17 (p<0.01)	0.26 (p<0.001)

Nota. Ambos ítems, el 24 y el «TBC-aciertos» miden el mismo hecho: la habilidad del alumno con el TBC. El ítem 24 de una forma subjetiva –es la opinión del alumno– y el «TBC-aciertos» de una forma objetiva, las puntuaciones que tuvo en el examen.

Solo el TBC mostraba estas correlaciones entre los enfoques de aprendizaje y la opinión positiva del alumno sobre el examen; el *test por aciertos* y el examen de *preguntas abiertas* prácticamente no tenían correlaciones significativas con los enfoques de aprendizaje. Esto nos hace creer que el uso del test con puntuación basada en el grado de seguridad puede inducir al alumno a adoptar enfoques más profundos de aprendizaje, lo que consideramos que sería una importante consecuencia positiva de este tipo de examen.

Pero estas correlaciones lo son entre ítems de una encuesta, reflejan solo las opiniones del alumno y, aunque lo sugieran, no prueban que el TBC premie el aprendizaje profundo y castigue el superficial. Tan solo la correlación entre los enfoques de aprendizaje profundos y el ítem 24, «*estuve acertado al escoger mi grado de seguridad*», indicaría algo más que la opinión del alumno, pues al hacer la encuesta este tenía toda la información sobre los resultados de sus exámenes y se le había proporcionado información exacta de su grado de acierto con el TBC. Es significativo que en este apartado concreto, acierto en el uso del TBC, más objetivo que los otros, la correlación más alta se dé con las estrategias y no con los motivos. Pero, con todo, la

respuesta del ítem 24 seguía siendo la opinión expresada por el alumno, que podía no ser sincero al contestar a esta pregunta.

Uno de los objetivos más importantes de nuestro estudio era el de ver si el TBC reconocía, y premiaba, los enfoques de aprendizaje profundos frente a los superficiales en comparación con los test convencionales. Esto lo hemos podido demostrar nosotros de forma fehaciente: **el TBC, a igualdad de aciertos en el test, premia con mayor puntuación a los alumnos que tienen enfoques profundos de aprendizaje**. Lo hemos demostrado, en comparación con el sistema de puntuación «*por aciertos*», para las estrategias profundas y, en menor medida, para los motivos profundos; pero no hemos probado que suceda lo contrario con los enfoques superficiales, que no influyen en la nota del TBC en mayor medida que lo hacen en la nota del test por aciertos.

Para poder verificar esto que hemos afirmado más arriba, necesitábamos comparar los enfoques de aprendizaje del estudiante, medidos según el RSPQ-2F (escala de Biggs), con datos objetivos, y no solo con las opiniones expresadas en la encuesta. Disponíamos de varios datos objetivos de cada encuestado: la nota obtenida por el sujeto en el TBC, la nota obtenida según número de aciertos y la diferencia entre ambas, que llamamos «TBC-aciertos». Estos datos eran exactos porque a los alumnos se les habían proporcionado, en un impreso aparte, esas puntuaciones y las tenían a la vista en el momento de rellenar la encuesta. Podían equivocarse o poner datos falsos, pero habíamos dispuesto un procedimiento para controlar que los datos que ponían eran los correctos. Comprobábamos que la pareja de cifras que habían puesto, puntuación en el TBC y número de aciertos, figuraban en el listado de notas del examen en el grupo (listado del que disponíamos con una clave en lugar del nombre del sujeto para respetar el anonimato); si un alumno modificaba una o las dos cifras, era extremadamente difícil

que volvieran a coincidir con alguna de las cifras del examen. Seguimos este procedimiento en los 270 estudiantes que proporcionaron el dato de su nota (2 no lo hicieron) y no nos encontramos con ningún caso en el que las notas aportadas en la encuesta no coincidieran con las notas reales. Asegurada así la exactitud de los datos de los resultados académicos de los estudiantes, sin vulnerar por ello el carácter anónimo de la encuesta, pudimos estudiar si el TBC reconocía, con mayor o menor puntuación, los enfoques de aprendizaje del alumno.

Primero analizamos la relación entre los enfoques y los resultados académicos. Hemos encontrado que los enfoques profundos de aprendizaje se relacionan con mejores resultados académicos mientras que los superficiales, en cambio, se correlacionan con peores resultados. Esto sucede tanto si el examen se corrige de la forma convencional, por aciertos, como si el examen es corregido según el grado de seguridad; pero con mayor correlación con el TBC en el caso de los enfoques profundos. Esto lo expresábamos en el capítulo dedicado a los resultados, que volvemos a poner a continuación, resumiendo las tablas 4.77, 4.78 y 4.79 en la tabla 5.9.

TABLA 5.9.
Correlación con los ítems de enfoques profundos y las puntuaciones del examen.

	Correlación con la nota en el TBC	Correlación con la nota en el test por aciertos
Enfoque profundo (29 +30 +33 +34 +37 +38 +41 +42 +45 +46)	0.250 (p<0.001)	0.195 (p<0.01)
Motivo profundo (29+33+37+41+45)	0.181 (p<0.01)	0.145 (p<0.05)
Estrategia profunda (30+34+38+42+46)	0.268 (p<0.001)	0.202 (p<0.001)

Nota. *r* de Pearson. En negrita el valor mayor de cada fila.

La correlación es siempre mayor con estrategias profundas que con motivos profundos, aunque las diferencias de correlaciones no alcancen significación estadística, lo que apoya la idea de que **el TBC refleja en mayor medida el aprendizaje profundo que el aprendizaje superficial**. No obstante, esa magnitud de correlaciones no era muy alta, para lo que esperábamos, pero separando los alumnos por años, resultó que los alumnos del año 2013 tenían una correlación importante (0.38 con enfoques profundos, y entre ellos 0.41 en estrategias) mientras que la correlación con los enfoques profundos en los alumnos de 2014 era de solo 0.12 (tabla 4.80). La razón estaba en uno de los dos grupos de Enfermería de 2014; un grupo de 55 alumnos cuyas respuestas en la escala de Biggs fueron en el sentido contrario a lo esperado por nosotros y a lo respondido en el resto de los grupos; este grupo de Enfermería tenía correlaciones inversas entre enfoques profundos y resultados académicos, lo que rebajaba la magnitud de los datos generales. Encontramos razones circunstanciales objetivas para el comportamiento anómalo de ese grupo, pero en aras de la honestidad del estudio, mantuvimos su inclusión, aceptando la reducción de correlaciones que ello suponía.

Para poder medir mejor el efecto de los enfoques profundos del aprendizaje sobre el tipo de examen, disponíamos del dato que nos daba el parámetro *nota con el TBC–nota por aciertos*, que llamamos «TBC–aciertos». Un *TBC–aciertos* positivo indicaba que el alumno se había conducido favorablemente con el sistema de examen basado en la confianza y había mejorado la nota que hubiera obtenido si el examen se hubiera corregido solo mediante aciertos; lo contrario indicaba, una escasa habilidad en el uso del TBC, indicaba un *TBC–aciertos* negativo. Si el test basado en la confianza premiaba a los alumnos que tenían enfoques de aprendizaje profundos habría una correlación positiva entre TBC–aciertos y enfoques profundos. Naturalmente, para poder valorarla hacía falta que la correlación tuviera suficiente magnitud como para ser

estadísticamente significativa. Dado que disponíamos de 270 sujetos, una correlación era significativa ($p < 0.05$) si su cuantía era igual o mayor que 0.1195.

El resultado, que hemos expuesto pormenorizadamente en tablas (4.89, 4.90 y 4.91) del capítulo anterior se expresa en 5.10 de forma unificada.

TABLA 5.10.

Correlaciones entre el resultado TBC–aciertos y enfoques, motivos y estrategias profundas.

	Enfoque profundo	Motivo profundo	Estrategia profunda
TBC–aciertos	0.195	0.128	0.229
Significación E.	Z = 3.19 ($p < 0.01$)	Z = 2.10 ($p < 0.05$)	Z = 3.75 ($p < 0.001$)

Nota. Resumida de las tablas 4.82, 4.83, 4.84 y 4.85. (r de Pearson)

Se demuestra una correlación positiva entre los enfoques de aprendizaje profundos y la habilidad en el uso del TBC, de manera que los alumnos con este tipo de enfoques mejoran la nota que obtendrían si hubieran sido evaluados solo mediante el número de aciertos. Esta correlación es mínima en cuanto a los motivos (0.13), pero es mayor en cuanto a las estrategias profundas de aprendizaje (0.23).

Vamos a destacar en esa tabla 5.11 los datos de la tercera columna de la tabla anterior, los de la correlación entre *estrategias profundas* de aprendizaje y la *habilidad en el uso del TBC*, o mejora de la nota que experimentan con el TBC frente a la puntuación por número de aciertos-que eso es lo que significa el índice *TBC–aciertos*-.

TABLA 5.11.

Correlación entre TBC-aciertos e ítems de estrategia profunda.

	Suma de ítems de estrategia profunda
Índice TBC–aciertos	0.229
Significación estadística	Z = 3.75 ($p < 0.001$)

El test basado en la confianza premia a los alumnos con enfoques profundos de aprendizaje, especialmente a los que desarrollan estrategias profundas.

Este resultado general se presenta con magnitud parecida en todos los grupos, de 2013 y de 2014, sin que aquí aparezca tan clara la distorsión que vimos habían causado las respuestas anómalas de uno de los grupos. La correlación entre *TBC–aciertos* con *estrategias profundas* es de **0.25** en el grupo de 2013 y de **0.21** en el de 2014, en ambos casos también significativas ($p < 0.01$ en el grupo de 2013 y $p < 0.05$ en el de 2014). Es necesario destacar que la nota *TBC–aciertos* no se correlaciona en absoluto con el nivel de conocimientos del alumno (expresado según el número de aciertos obtenido en el examen); ya habíamos visto que los enfoques profundos de aprendizaje se correlacionaban con los resultados académicos, por lo que si hubiera una correlación positiva y potente entre *TBC–aciertos* y número de aciertos, la correlación entre enfoques profundos y *TBC–aciertos* podría ser dependiente de esa otra, pues los alumnos con enfoques profundos de aprendizaje obtienen puntuaciones más altas. Pero entre el *TBC–aciertos* y la nota por aciertos hay una correlación insignificante y, además, de signo negativo (r de -0.05); es decir, la habilidad en el uso del TBC es igual en los alumnos con pocos conocimientos que en los que tienen muchos. Así que la correlación positiva entre la habilidad en el uso del TBC, medida mediante el índice *TBC–aciertos*, y las estrategias profundas es independiente de cualquier otro factor detectable en la encuesta.

La magnitud de esta correlación no es muy alta, una r de 0.23 se podría considerar una correlación débil, pero es muy consistente, pues se repite en todos los años y en todos los grupos, y es también muy significativa, con una Z de 3.75 y $p < 0.001$.

Los factores que influyen en el resultado de un examen de tipo test, además del principal, que es el nivel de conocimientos del alumno, son múltiples; en un examen de tipo TBC son aún más, pues en lugar de un solo componente de la nota, el número de aciertos, hay tres: aciertos, autoconfianza y adecuación en cada pregunta del acierto y la autoconfianza. Influye, naturalmente, la suerte (minimizada cuando hay un alto número de preguntas) y el cálculo estratégico de alumno, que puede no ser sincero al expresar su opción de confianza buscando un objetivo de puntuación. Así mismo, la concurrencia de una gran cantidad de factores complejos que influyen en el sujeto al tomar una decisión, y en este tipo de test el estudiante toma dos decisiones simultáneas en cada pregunta, hace que poder probarla relación entre unas estrategias utilizadas meses, semanas o días antes del examen y una modificación objetiva pero sutil en la puntuación del examen (la diferencia entre la nota TBC y la nota por aciertos suele ser de unas pocas décimas de punto, en el 67 % de los casos inferior a 0.6 puntos) se presente como sumamente difícil. En esas condiciones, haber conseguido probar que las estrategias profundas de aprendizaje influyen en la habilidad del alumno para escoger su nivel de confianza, mejorando en consecuencia su nota en la evaluación, consideramos que es relevante y meritorio; en nuestra opinión, es un hallazgo conspicuo del presente estudio.

Este hallazgo hay que ponerlo en relación con lo que vimos que ocurría con el ítem 24, en el que se le preguntaba al alumno sobre su habilidad al escoger su opción de confianza -«*¿has estado afortunado al escoger tu nivel de seguridad?*»-. Aunque en esta pregunta el alumno respondía con su libre opinión, esta era más objetiva que las opiniones pedidas por los demás ítems, pues al estudiante se le había proporcionado información sobre su resultado en el examen y su habilidad en el uso de TBC. Este ítem también se correlaciona positivamente con los enfoques profundos, e, igualmente, más

con las estrategias profundas que con los motivos profundos: r de 0.24 con enfoques en general, 0.17 con motivos y 0.26 con las estrategias profundas (tabla 5.10). A destacar el muy notable paralelismo con las correlaciones con TBC-aciertos-0.20, 0.13 y 0.23 respectivamente- como expusimos en la tabla 5.8.

Tabla 5.8.

Paralelismo en las correlaciones de los enfoques profundos de aprendizaje con la habilidad objetiva y la autopercibida en el uso del TBC (TBC-aciertos/ítem 24)

	Enfoques profundos	Motivos profundos	Estrategias profundas
TBC-aciertos (habilidad objetiva)	0.20 ($p < 0.01$)	0.13 ($p < 0.05$)	0.23 ($p < 0.001$)
Ítem 24 (habilidad subjetiva)	0.24 ($p < 0.001$)	0.17 ($p < 0.01$)	0.26 ($p < 0.001$)

Nota. r de Pearson.

También complementa este hallazgo -la relación entre la habilidad para escoger la opción de confianza y los enfoques profundos de aprendizaje- los otros datos de la encuesta que informaban de los efectos del TBC sobre el estudio. Los alumnos que opinan en mayor medida que el TBC les hace estudiar más y les hace comprender mejor eran los que tenían enfoques profundos de aprendizaje (tabla 5.5); no resulta raro que esos alumnos, que dedican más tiempo a preparar los exámenes e intentan comprender mejor lo que aprenden, sean los que tienen motivos y estrategias más profundas de aprendizaje y los que escogen mejor su opción de seguridad. El mayor tiempo dedicado al estudio y, especialmente, el mayor empeño en comprender lo que se aprende, no solo aumentarían el conocimiento, sino que profundizarían la metacognición del estudiante, y así, no solo sabría más sino que conocería mejor lo que sabe y lo que ignora; su dominio de la materia sería, por consiguiente, superior. **El TBC, al premiar estas actitudes de mayor esfuerzo de comprensión y estrategias más profundas en el estudio, podría potenciar un mayor y mejor aprendizaje que el test convencional por aciertos.** La propia opinión de los alumnos va a favor de estas ideas, como queda

expresado cuando opinan que el TBC es el examen que «les hace aprender más» (ítem 22, TBC obtiene una puntuación de 5.07 frente a un 4.48 de examen de preguntas abiertas y un 4.38 del test por aciertos); los que opinan eso son los que precisamente tienen enfoques más profundos. Así, **los datos objetivos recogidos al estudiar la correlación de la nota TBC-aciertos y los enfoques profundos de aprendizaje confirman y complementan los subjetivos obtenidos en la encuesta.**

A diferencia de lo que ocurre con los enfoques profundos, especialmente las estrategias profundas, en los enfoques superficiales de aprendizaje no aparece una correlación significativa con la habilidad en el TBC. El coeficiente r es muy pequeño, -0.03 para enfoques superficiales y -0.06 para las estrategias, y carece de toda significación. El TBC, por tanto, no penaliza los enfoques superficiales de aprendizaje.

Así que, en resumen, el **test basado en el grado de confianza premia las estrategias profundas de aprendizaje y penaliza a los alumnos que no las utilizan**, en mayor medida que el test convencional. No penaliza, en cambio, las estrategias superficiales más de lo que lo hace el test por aciertos.

5.7. CONCLUSIONES DEL ESTUDIO DE LOS ÍTEMS DE AUTORREGULACIÓN Y SUS CORRELACIONES

Debido a nuestro deseo en no sobrecargar a los estudiantes con una encuesta excesivamente larga nos impusimos una limitación en el número de ítems -no queríamos que fueran más de 50, al final fueron 52-. Como la sub-encuesta sobre la opinión del alumno respecto a los exámenes ocupaba 27 ítems y la escala RSPQ-2F otros 20, no pudimos poner un cuestionario completo para valorar la autorregulación del

alumno; el «cuestionario de aprendizaje autorregulado» de Juan Carlos Torre (2006) tiene 20 ítems. Pero decidimos incluir algunas preguntas de dicho cuestionario en número pequeño para hacer un estudio preliminar exploratorio que pudiera orientar posteriores investigaciones, así como completar la visión que los enfoques de aprendizaje nos daban. Escogimos 5 preguntas en total, eligiendo las que nos parecieron más apropiadas, basándonos en una encuesta parcial previa hecha en 2012 (datos no publicados); nuestro objetivo fue disponer de un agrupamiento que tuviera una consistencia superior o, al menos, próxima a 0.70 (media por el índice α de Cronbach). Esas cinco preguntas no pertenecían únicamente a uno de los cuatro factores del cuestionario, sino que se repartían entre dos factores: *conciencia metacognitiva activa*, ítems 50 y 51, y *control y verificación*, ítems 49, 52 y 53. Este agrupamiento nos proporcionó en el presente trabajo un valor de consistencia interna (alfa de Cronbach) de 0.735, bastante bueno para una encuesta de solo 5 ítems. Las puntuaciones medias otorgadas a estas preguntas por los encuestados fueron elevadas, con una media de 4.56 (de 1 a 6) –lo que supone el 71.2 % de la puntuación máxima–; con importantes diferencias entre los ítems, que van de 3.67 en el ítem 51 a 5.10 en el 53 –55.4 % a 82.0 %–. Lo que nos interesó más de este agrupamiento fueron sus correlaciones.

Las puntuaciones altas en autorregulación se correlacionan significativamente con los siguientes ítems o factores:

- Con los buenos resultados en el examen; r de 0.344 con la nota en el TBC y 0.310 con la del test por aciertos.
- Con la buena elección del nivel de confianza (TBC–aciertos); r de 0.148.

- Con la opinión de haber escogido bien el nivel de seguridad (ítem 24), r de 0.291, y con al autoeficacia docente, r de 0.273.

- Con los enfoques de aprendizaje, tanto con los profundos –correlación positiva– como con los superficiales –correlación negativa en este caso–; estas correlaciones son las más elevadas del factor *autorregulación*. Con los enfoques profundos la r de Pearson es de 0.471 (0.40 con *motivo profundo* y 0.45 con *estrategia profunda*). Con los enfoques superficiales la correlación es negativa, de –0.331 (0.31 *motivo superficial* y 0.28 *estrategia superficial*). Los alumnos con enfoques profundos tienen un aprendizaje autorregulado mayor, y los que tienen enfoques superficiales, menor.

- Con tener una opinión positiva sobre el TBC (y no con los demás tipos de examen). Con el ítem 3, que expresa predilección por el TBC, la correlación es pequeña (0.134), con el ítem 22, ya visto, era de 0.288 y con el que expresa que el alumno es hábil utilizando el TBC la r es de 0.291. Con el agrupamiento de tres ítems que denominamos «validez del TBC» la correlación es de 0.195 y con el agrupamiento «eficacia docente» las opiniones positivas sobre el TBC en todos los sentidos la r es de 0.233. Los alumnos con mejor autorregulación tienden a tener una opinión positiva sobre el TBC, algo que no ocurre con los demás tipos de examen.

De acuerdo con todo lo anterior, los estudiantes con mejor aprendizaje autorregulado, en la medida un tanto limitada en que lo hemos estudiado, tienen, por término medio, mayor puntuación en los enfoques de aprendizaje profundos y menor en los superficiales. Obtienen mejores resultados académicos y tienen una opinión mejor

del TBC, tanto en lo que respecta a su validez como sistema de examen como en lo que respecta a su experiencia personal con él. Escogen ligeramente mejor la opción de seguridad y tienen mayor autoeficacia docente.

5.8. INFLUENCIA DEL GÉNERO DE LOS ESTUDIANTES EN LOS RESULTADOS EN EL TBC

En nuestro estudio fueron incluidos 271 estudiantes que cumplimentaron el dato del género; 216 era mujeres y 55 varones. Hemos encontrado muy pocas correlaciones de los demás ítems de la encuesta con el género, solo cuatro:

- Las estudiantes mujeres dan puntuación más alta a la pregunta que interroga sobre el estrés que genera cada tipo de examen (ítems 26, 27 y 28). Los tres tipos generan más estrés en las mujeres (r de 0.34 en el caso del test por aciertos, de 0.32 en el TBC y de 0.24 en el de preguntas abiertas). Las puntuaciones de cada género las hemos expuesto en las tablas 4.68, 4.69 y 4.70 de resultados. En resumen, son las siguientes:
 - **Estrés con el test por aciertos:** mujeres 4.62 (δ 1.296), hombres 3.42 (δ 1.583), $F=5.853$, $p<0.001$, $d=0.90$.
 - **Estrés con el TBC:** mujeres 5.08 (δ 1.112), hombres 4.05 (δ 1.649), $F=18,96$, $p<0.001$, $d=0.83$.
 - **Estrés con las preguntas abiertas:** mujeres 5.20 (δ 1.173), hombres 4.44 (δ 1.561), $F=14.28$, $p<0.001$, $d=0.60$.

- La cuarta correlación significativa es la relación del género con los *motivos superficiales* de aprendizaje. En este caso la correlación es positiva con los varones y su magnitud es débil: r de 0.15 ($p < 0.05$).

Más destacable que las correlaciones existentes es la ausencia de otras correlaciones. Así, en el nuestro estudio global el género no se correlaciona significativamente con los resultados académicos (r de Pearson de 0.02) ni con la habilidad en el uso de TBC (medida por el índice *TBC-aciertos*, r de Pearson de 0.01). Tampoco hay correlación significativa con la opinión sobre los distintos tipos de examen, con la autoconfianza, con enfoques profundos de aprendizaje ni con la autorregulación.

Podemos afirmar que en nuestro estudio no aparece sesgo de género en ningún sentido relacionado con la habilidad en el desempeño del estudiante con el TBC. El TBC no tiene, por lo tanto, sesgo de género.

5.9. CONCLUSIONES DEL ESTUDIO DE LA AUTOCONFIANZA DE LOS ALUMNOS

El TBC nos proporciona información no solamente sobre los conocimientos que el estudiante tiene de la materia, sino también sobre su autoconfianza. En teoría, debería haber un paralelismo entre los conocimientos y la confianza, pues esta se basa en la noción que el sujeto tiene de aquellos; dicho de otra manera, el alumno que sabe mucho cabe pensar que tenga un alto nivel de confianza en el acierto de sus respuestas, y el que tiene pocos conocimientos tendrá un nivel de confianza más bajo. En la práctica esto se da, más o menos, pero con numerosas desigualdades y excepciones. Al corregir los

exámenes encontramos estudiantes con pocos aciertos y mucha seguridad, y otros, en cambio, con un alto porcentaje de aciertos pero muy inseguros –alumnos «sobreconfiados» e «infraconfiados» en relación a sus conocimientos-. Se supone que el sistema premia al alumno cuando su seguridad o confianza es pareja a sus conocimientos, y castiga tanto al exceso temerario como al defecto timorato de confianza (Gardner–Medwin, 2006). Esta afirmación es algo que se deduce lógicamente de las tablas de puntuación de los exámenes basados en la confianza, que premian el acierto con seguridad y penalizan el error confiado; pero eso es solo teoría y no estaba verificado en la práctica. La cuestión es: *la afirmación de que la puntuación es máxima cuando la confianza es pareja al conocimiento ¿se cumple en la práctica?* Nosotros queríamos comprobar que esto era así efectivamente, y la medida en que era así, por lo que diseñamos un instrumento para medir la confianza del alumno con independencia de sus respuestas correctas e incorrectas. Con ese instrumento pudimos comprobar si el grado de confianza medio era parejo a los conocimientos y si la discrepancia entre ambos suponía, como creíamos que debía ser, una penalización en la nota.

El instrumento para medir el grado de confianza lo diseñamos en nuestro trabajo preliminar y lo hemos explicado en los capítulos «Marco teórico» y «Metodología». Al instrumento en cuestión lo llamamos *puntuación en confianza*, PC en siglas. Como explicamos, este parámetro surge de asignar una puntuación máxima (10) a las preguntas en las que el examinando ha optado por la máxima seguridad, y la mitad de esa puntuación (5) a aquellas en las que escoge la opción intermedia, dando cero a la opción de confianza mínima. En el capítulo «Estado de la cuestión» tuvimos ocasión de explicar las bases racionales de esta asignación de puntuación y de comprobar lo adecuada que era. Otros autores, como mencionamos en el Marco Teórico, han estudiado también el grado de confianza del estudiante en el examen, utilizando escalas

que lo medían con independencia de las notas obtenidas en la evaluación (Morony y otros, 2013). Los autores, encontraron, igual que nosotros, una alta correlación entre la confianza del estudiante, medida durante el examen y referida a la probabilidad de acertar o fallar cada cuestión, y el resultado real de la evaluación

La PC nos proporciona una ajustada medida del grado de seguridad del alumno. A partir de ella y del número de aciertos obtenido por el alumno, obtenemos otros tres parámetros más:

- El cociente $PC/nota\ por\ aciertos$, al que hemos llamado *índice de confianza* (IC). Este parámetro es > 1 si el sujeto tiene más confianza que aciertos y < 1 en la situación opuesta.
- El parámetro $PC-aciertos$, que mide la diferencia entre el PC y la *nota por aciertos*, con signo positivo -mayor de 0- cuando el estudiante tiene más confianza que conocimientos o negativo -menor de 0- en el caso contrario.
- El parámetro $|PC-aciertos|$, que es el parámetro anterior pero en valor absoluto, sin signo negativo, nos mide la dispersión de la autoconfianza del alumno con respecto al valor teórico ideal, que sería de 0; no tiene utilidad individual, sino en estudios del comportamiento del grupo, midiendo la dispersión respecto a la *paridad confianza-conocimientos* de los estudiantes.

Los dos primeros parámetros, IC y $PC-aciertos$, nos permiten conocer hasta qué punto el alumno escoge asumir un riesgo excesivo, un riesgo adecuado o un riesgo escaso. Se pueden obtener directamente al corregir el examen, incluso si se corrige usando un sistema de lectura automática, y nos darán información sobre la

autoconfianza y el riesgo asumido por el alumno en el examen, lo que es útil tanto para la actividad de tutorías docentes como para investigaciones. Son cifras que, añadidas a su puntuación, nos proporcionan abundante información sin necesidad de buscar y recuperar el examen del alumno. Además, permiten seguir la evolución del alumno en el tiempo, proporcionando una dimensión suplementaria a los resultados académicos.

El IC, índice de confianza, que se obtiene del cociente $PC/nota\ por\ aciertos$, es un valor numérico que presenta una distorsión, que llega a ser importante, cuando la puntuación del alumno es baja. Por ejemplo, en un alumno que tuviese de *nota por aciertos* 0, el IC podría ser infinito, impidiendo cualquier cálculo de los valores del grupo. Aparte de esa posibilidad, que estuvo a punto de suceder en un caso, la estructura matemática del IC tiende a aumentar los valores del índice en los alumnos con menor puntuación y a disminuirlos en los que obtienen una puntuación más alta. Por ello, hemos decidido prescindir de este indicador en el futuro, aunque en el presente estudio lo hemos calculado para compararlo con nuestros datos del trabajo preliminar realizado en 2009 y 2010.

Para analizar estos datos solo hemos dispuesto de los estudiantes del año 2014, pues en la encuesta de 2013 se omitió el dato de la puntuación en confianza. En total han sido 139 sujetos los estudiados (108 mujeres, 31 varones).

La PC media resultó extraordinariamente próxima a la media de la nota de los estudiantes. Esta proximidad la expusimos en la Tabla 4.139, referida a los datos del presente trabajo de 2014. En ella mostramos que la nota media obtenida por los alumnos fue de 6.16 con el TBC, 6.20 con el test *por aciertos* y 6.11 en *puntuación en confianza* –PC-. Solo 5 o 9 centésimas de punto de diferencia entre una nota real, obtenida en el examen, y la autoconfianza media de los alumnos.

Resulta llamativa la casi igualdad entre las notas reales obtenidas en el examen y la *puntuación en confianza* medida según los parámetros que hemos establecido; hay que tener en cuenta que la PC proporciona una medida de la opción de confianza del estudiante con independencia de que las respuestas a las preguntas sean correctas o incorrectas. En nuestro estudio preliminar, realizado con los resultados de 2008 y 2009, obtuvimos también una gran paridad entre la PC y las notas de los exámenes (tabla 3.15 de Metodología); ahora, en los exámenes realizados en 2014 revalidamos esos resultados: la puntuación media en confianza (según nuestro sistema de calcularla) es casi igual a la nota media obtenida en el examen por los alumnos.

Por término medio los alumnos poseen una autoconfianza equivalente a sus conocimientos, lo que resulta un hallazgo positivo que refuerza la opinión de que el TBC es un sistema de evaluación consistente y que la PC que hemos establecido es un buen indicador de la confianza real del estudiante. Pero el que la nota media del examen y la puntuación media en confianza de los alumnos sean casi iguales no quiere decir que los alumnos tengan siempre valores muy parecidos en ambos parámetros. De hecho, las cifras de IC y *PC-aciertos* nos muestran que, aunque la media de la nota y de la puntuación en confianza en el grupo total es muy próxima, hay grandes diferencias individuales. Así, el rango de puntuaciones en *PC-aciertos* es de 8.93, esa es la suma de diferencias entre los dos estudiantes con mayor diferencia entre conocimientos y confianza en sentido positivo (+5.35) y negativo (-3.58) (datos recogidos en la tabla 4.139). Hay alumnos demasiado cautelosos y otros, en cambio, son temerarios, pero la mayoría tienen un nivel de confianza muy ajustado a su nivel de dominio de la materia y la puntuación en confianza media del grupo solo se separa en unas centésimas de punto de la nota media en el examen.

La medida de la dispersión media de la diferencia entre confianza y conocimientos de los estudiantes sobre el valor medio próximo a 0, expresada en cierta manera en la desviación estándar de los índices IC y *PC-aciertos*, se percibe con más claridad en los estadísticos del parámetro que hemos llamado $|PC-aciertos|$ (*PC-aciertos* en valor absoluto), que mide esta diferencia sin tener en cuenta el signo. Dichos estadísticos los expusimos en la Tabla 4.139, que ahora repetimos resumida

TABLA 5.12.
Estadísticos de $|PC-aciertos|$

	$ PC-aciertos $
Media	1.11
Desviación típica	0.928
Mediana	0.87
Mínimo-máximo (rango)	5.35

Nota. Datos extraídos de la tabla 4.139.

En ella podemos ver que la media de las diferencias entre la *nota por aciertos* y la *puntuación en confianza* es de 1.11 y la mediana de 0.87. El 50 % de los estudiantes tienen una diferencia superior a 0.87 (valor de la mediana) y el 33 % superior a 0.93 (valor de la desviación estándar); la diferencia entre confianza y conocimiento llega a ser en un caso de 5.35 sobre 10. Por consiguiente, aunque la media de la puntuación en confianza y la puntuación según aciertos es casi la misma, los valores individuales de la diferencia de ambos parámetros muestran una gran dispersión; como hemos expresado en la tabla 4.138 del capítulo «Resultados». Hay sujetos que exhiben una gran autoconfianza y escasos conocimientos (hasta 5.35 puntos sobre 10 de diferencia), y otros, en cambio, grandes conocimientos pero con una autoconfianza pequeña (hasta – 3.58 puntos sobre 10); el 50 % de los sujetos estudiados tienen una discrepancia entre conocimientos y confianza menor de 0.87 puntos.

Mediante el TBC, utilizando la *puntuación en confianza* y la comparación de esta con la nota realmente obtenida (índice *PC-aciertos*), podemos evaluar si los conocimientos del alumno, establecidos por su nota en el examen, son parejos a su *confianza en los conocimientos* (Gardner-Medwin, 2006), establecida por el PC. Esto complementa y enriquece la información que proporciona el proceso evaluador; nos hace conocer un rasgo de la personalidad del sujeto, su «tendencia a arriesgarse», relevante en las ciencias de la salud, y nos ofrece datos para orientar su tutoría, hacerle ver sus defectos y corregir sus debilidades.

La discrepancia entre el conocimiento y la confianza del alumno se puede deber a un déficit en la metacognición de éste –que realmente no sepa hasta dónde llegan sus conocimientos y sus lagunas cognitivas-, a una falta de sinceridad sobre su auténtico nivel de seguridad, motivada por un cálculo estratégico sobre su nota, o a una tendencia psicológica mayor o menor de lo habitual a sumir riesgos o a rehuirlos.

Para poder conocer en qué medida influye cada uno de esos factores, estratégico, déficit de metacognición y tendencia a arriesgarse, en la elección por los estudiantes de su nivel de seguridad, disponemos de algunos indicios, aunque los parámetros de la evaluación no nos permiten efectuar deducciones con total seguridad. Uno de esos indicios surge del estudio de las correlaciones de la encuesta, tal como hemos expuesto en la tabla 4.140 del capítulo «Resultados». La reproducimos a continuación.

TABLA 4.140.

Correlaciones entre la autoconfianza y las notas del examen (r de Pearson).

	<i>PC</i>	<i>IC</i>	<i>PC-aciertos</i>	<i>PC-aciertos</i>
Nota TBC	0.749	-0.471	-0.404	-0.215
Nota por aciertos	0.681	-0.549	-0.549	-0.086
TBC-aciertos	0.216	0.236	0.447	-0.401

Hay una correlación negativa importante entre el IC y *PC-aciertos*, por un lado, y la nota obtenida por el alumno, r de -0.549 con ambos índices. Eso indica que los alumnos con nota menor tienden a asumir, proporcionalmente, más riesgos que los alumnos con nota más elevada. Dichos alumnos con nota baja, para tener esperanza de aprobar el examen, no pueden ser sinceros en la expresión de su auténtica confianza, pues los aciertos con baja seguridad proporcionan solo un tercio de la nota posible, insuficiente para aprobar, sino que han de arriesgarse manifestando una confianza mayor que la real, en la esperanza de tener suerte.

Por otro lado, los alumnos con buen nivel de conocimientos pero que no ambicionan obtener las notas máximas (sobresaliente o matrícula de honor) y que aspiran simplemente a asegurar su aprobado, tienden con frecuencia a no asumir riesgos. Esta relación inversa entre riesgo relativo y los conocimientos, que el estudio de correlaciones nos muestra, es algo que constatamos personalmente al corregir los exámenes con este sistema. Reiteradamente hemos visto cómo los alumnos de la franja de notas entre 6 y 8, con aprobado seguro pero demasiado alejados del sobresaliente, tienen una excesiva prudencia que les acaba perjudicando en su puntuación, mientras que los casos de mayor riesgo, rayano en la imprudencia académica, se dan entre los alumnos con menores conocimientos. La mayor discrepancia en sentido positivo entre la autoconfianza y la nota real por aciertos se ha dado en una alumna de Enfermería (*sujeto 27 de 2.ºB, 2014*) que obteniendo una nota *por aciertos* de 4.05 tiene una puntuación en confianza de 9.4; es decir, tiene conocimientos de «suspense» pero declara tenerlos de «sobresaliente». La mayor en sentido negativo se da en un alumno de Fisioterapia (*sujeto 37 de Fisioterapia, 2014*) que con una nota *por aciertos* de 9 tiene una puntuación en confianza de solo 5.4; es decir, teniendo conocimientos de sobresaliente escoge una opción de seguridad que corresponde a un aprobado escaso.

Ambos alumnos salieron bastante perjudicados en su nota por la discrepancia entre conocimientos y seguridad: la primera estudiante obtuvo un 2.80 con el TBC, frente al 4.05 que hubiera tenido corrigiendo su examen por aciertos, y el segundo obtuvo un 7.80 con el TBC frente a un 9 que hubiese tenido en un examen por aciertos. En ambos casos el sistema de evaluación basado en la confianza les penalizó restándoles 1.20 puntos por su falta de sinceridad o por su inseguridad.

El TBC tiende a penalizar las discrepancias entre el conocimiento declarado (mediante la opción de seguridad) y el conocimiento real. Esto lo constatamos en los resultados de las correlaciones de índice $|PC\text{-aciertos}|$, que miden esa discrepancia sin tener en cuenta el signo. Analizando la correlación entre la diferencia $TBC\text{-nota por aciertos}$, que utilizamos para medir la habilidad del alumno con el TBC, y el índice $|PC\text{-aciertos}|$ obtuvimos los siguientes resultados:

TABLA 5.13.
Correlaciones $|PC\text{-aciertos}|$ – notas del examen.

	$ PC\text{-aciertos} $
Nota por aciertos	NS
TBC–aciertos	-0.401 p < 0.0001

Nota. Tabla resumida de la 4.140 de «Resultados»

La correlación negativa destacada de -0.401 indica que cuanto más se separa la confianza del alumno de sus conocimientos reales, menos puntuación obtiene con el TBC, cosa que no ocurre con el test por aciertos. EL TBC penaliza, por consiguiente, la falta de sinceridad y/o de metacognición del alumno.

Podríamos suponer que un índice $|PC\text{-aciertos}|$ próximo a cero, que indica un buen ajuste entre las expectativas y los resultados, en orden a acertar o fallar las

preguntas, debería ir asociado a un mayor número de aciertos (Nicholson, Putwain, Connors y otros, 2013), pues los alumnos mejor formados suelen tener expectativas más realistas, pero nosotros no hemos podido constatar ese hecho, como se aprecia en el hallazgo de que no hay correlación significativa entre el índice $|PC\text{-aciertos}|$ y el número de aciertos (r de 0.086, $p > 0.05$).

La magnitud del premio o castigo que el TBC otorga por los motivos antes expuestos no es muy elevada, la desviación estándar del valor $TBC\text{-aciertos}$ es de 0.63 puntos, pero es reconocida por el alumno cuando, tras el examen, se le retroalimenta con la información de la nota que obtuvo y la que hubiera obtenido si hubiera estado más atinado al escoger sus opciones de confianza, cosa que se debería hacer siempre con este sistema de examen para obtener de él toda su potencialidad, y que nosotros hacemos sistemáticamente.

5.10. RELACIÓN ENTRE GÉNERO Y AUTOCONFIANZA

La capacidad del TBC para analizar el grado de confianza de los alumnos nos permite estudiar si el género tiene relación con la tendencia al riesgo o la prudencia en los exámenes. Nuestros resultados en el trabajo preliminar, no habían encontrado diferencias significativas entre los dos sexos a este respecto, pero habiendo sido obtenidos con un número limitado de alumnos, 88 entre hombres y mujeres, tenían una relevancia limitada (tabla 4.141). Disponiendo del IC y del $PC\text{-aciertos}$ como herramientas para medir la tendencia a asumir riesgos, hemos vuelto a hacer este análisis en los estudiantes incluidos en el presente trabajo en el año 2014, para ver si esos datos se confirmaban o no. Hemos dispuesto esta vez de un número algo mayor de

sujetos en los que disponíamos del dato del género y la puntuación en confianza, 139 en total, pero con desproporción entre ambos sexos pues solamente 31 de ellos eran varones. Con estas limitaciones consideramos que sería interesante repetir el estudio para ver si se confirmaba o se desmentía al anterior, realizado en 2009. Los resultados del presente trabajo mostraron de nuevo que no hay diferencia significativa entre nuestros estudiantes varones y mujeres en cuanto a su tendencia a asumir riesgos en el examen. Para analizar estos datos hay que tener en cuenta que en el grupo analizado, correspondiente a los estudiantes de 2014, las mujeres obtuvieron unos resultados académicos ligeramente superiores a los varones (6.2 frente a 5.9 de media), lo que hacía que su puntuación en confianza fuera algo superior a la de ellos, por la relación ya vista entre resultados académicos y confianza (PC de 6.2 en mujeres frente a 5.6 en varones). El índice *PC-aciertos* neutraliza esas diferencias en los resultados académicos, y con él hemos encontrado que no hay diferencias significativas entre varones y mujeres en cuanto a su tendencia a asumir riesgos en nuestros exámenes. En las tablas 4.128 y 4.130 de «Resultados» hemos expuesto los resultados de nuestros estudios sobre la relación del género con la tendencia a arriesgarse en el TBC. Habiendo pasado 5 años entre el estudio preliminar (realizado en alumnos de 3º de Enfermería y Fisioterapia en 2009) y el actual (realizado en alumnos de 2º de Enfermería y Fisioterapia en 2014) consideramos que hay evidencias, sostenidas en el tiempo, para afirmar que entre nuestros estudiantes el género no parece tener influencia significativa en cuanto a su tendencia a la prudencia o al riesgo al escoger su opción de seguridad en el TBC. Reproducimos las tablas expuestas en el capítulo anterior, para la PC-aciertos, la 4.143.

TABLA 4.143.

Datos estadísticos de la PC-aciertos

sexo	N	PC-aciertos	δ	Error T.	d	F	t	p
mujer	108	-0.22	1.426	0.138	0.10	0.047	0.485	>0.05
varón	31	-0.08	1.482	0.266				NS

Nota. Estudio realizado en los alumnos de 2º de Enfermería y Fisioterapia en febrero-marzo de 2014

La diferencia es muy pequeña y no significativa. Aunque de unos datos que no tienen significación estadística no deberían extraerse conclusiones, nos llama la atención que en nuestro estudio la tendencia al riesgo salga ligeramente superior en las mujeres que en los hombres. En cualquier caso sin significación.

Al estudiarlos datos de 2014 mediante un procedimiento alternativo, que consiste en ver las correlaciones entre el género y los parámetros que miden la confianza, obtenemos, como debe ser, los mismos resultados: no hay correlación significativa ni con el IC, r de 0.015, ni con el índice *PC-aciertos*, r de 0.041.

Podemos así afirmar que, al igual que nos sucedió en nuestro estudio anterior, no encontramos que haya ninguna relación entre el género de nuestros estudiantes y su tendencia asumir riesgos en el examen. Estos resultados difieren de los encontrados en otros estudios sobre la relación entre tendencia al riesgo y género, como explicaremos en el capítulo siguiente, «Discusión».

5.11. CONCLUSIONES RELEVANTES RESPECTO A LOS OTROS DOS SISTEMAS DE EXAMEN

Aunque el objetivo de nuestro trabajo es el de analizar el TBC, en nuestra encuesta hemos incluido ítems que interrogaban al alumno sobre los test convencionales por número de aciertos y sobre los exámenes de preguntas abiertas, utilizándolos como

patrones con los que comparar el TBC. De las respuestas de los estudiantes a las preguntas sobre esos otros dos tipos de examen hemos obtenido las siguientes conclusiones:

Con el examen de tipo *test basado en el número de aciertos*:

- Es el tipo de examen, de los tres incluidos en el estudio, que los alumnos prefieren.
- Es el examen que les obliga a dedicar menos tiempo al estudio.
- Es el examen que les obliga a hacer un menor esfuerzo de memorización.
- Les obliga a hacer un menor esfuerzo de comprensión que el TBC, aunque igual que el examen de preguntas abiertas.
- Los alumnos opinan que es el examen de menor «validez autopercebida» (subfactor formado por tres preguntas: sobre la información que el examen al profesor de lo que el alumno sabe –ítem 15-, la utilidad para la autoevaluación del alumno –ítem 18- y el estímulo general para el aprendizaje –ítem 21-).
- Es, en opinión de los estudiantes, el examen con menor *eficacia docente*. El factor «eficacia docente», lo habíamos formado con el agrupamiento de cinco preguntas, las tres del agrupamiento del párrafo anterior y dos que hacen referencia a las modificaciones en el estudio inducidas por cada examen, aumento del tiempo dedicado al estudio –ítem 5 en este caso- y del esfuerzo de comprensión –ítem 8-.

Con el examen de preguntas abiertas:

- Es el examen que al alumno le provoca más aversión, el que menos le gusta.
- Es el examen que le obliga a dedicar más tiempo al estudio.
- Es el examen que le obliga a hacer un mayor esfuerzo de memorización.
- Los alumnos consideran que es un examen con mayor *validez autopercebida* y *eficacia docente* que el test por aciertos, pero peor que el TBC en ambos conceptos.
- Los alumnos que consideran que el examen de preguntas abiertas constituye un buen sistema de examen que ayuda al aprendizaje –*eficacia docente*-, -los que otorgan a este examen altas puntuaciones en los ítems 7, 10, 17, 20 y 23- constituyen un grupo muy bien definido y homogéneo, con una consistencia interna de 0.891 (alfa de Cronbach).

En los tres sistemas de examen que hemos analizado en la encuesta se encuentra una importante asociación entre considerar que un tipo de examen obliga a entender bien lo que se estudia y tener la opinión de que es un buen sistema de evaluación, con una alta eficacia docente, -la correlación media de estos dos conceptos en los tres exámenes es de 0.537-. Los alumnos consideran que el buen aprendizaje debe fomentar la comprensión. Por ese motivo, el que nuestros estudiantes consideren que el TBC es el que fomenta más la comprensión, por encima del test convencional y del examen de

preguntas abiertas, es tan relevante e indica que, aunque no sea el procedimiento de evaluación que más les gusta, reconocen que es, simplemente, el mejor.

6. DISCUSIÓN Y CONCLUSIÓN FINAL

6.1. CARACTERÍSTICAS DE UN BUEN SISTEMA DE EVALUACIÓN

La evaluación fue por primera vez colocada en la cúspide del proceso de aprendizaje en la taxonomía original de Bloom (1956), y en sus derivaciones posteriores (Anderson y Krathwohl, 2001) ha mantenido esa posición cimera. Pero a medida que el aprendizaje se ha ido volviendo más complejo, más rico y potente, la evaluación se ha vuelto más complicada, aunque también más importante (Urosa, 1995). Si evaluar es «hacer juicios», como dice Mayer (2002), estos juicios se han vuelto, con el tiempo, muy variados en sus procedimientos, y cada vez más complejos en su interpretación.

El profesor dispone de un arsenal completo de procedimientos de evaluación: exámenes de diversos tipos –preguntas abiertas largas o cortas, test de múltiples formatos...-, problemas, casos-problema, redacciones, trabajos, simulaciones prácticas,... etc., y con frecuencia se encuentra con dudas a la hora de elegir el procedimiento mejor. Nosotros creemos que es recomendable, siempre que las condiciones objetivas del medio docente lo permitan, una combinación de varias técnicas de evaluación, para analizar el aprendizaje desde puntos de vista diferentes y complementarios. Pero sea cual sea el procedimiento escogido, la técnica evaluativa ha de cumplir unos criterios o estándares previos. Esos criterios Richard Mayer los definió como *calidad, consistencia, eficiencia y efectividad* (Mayer, 2002).

La *calidad* de un examen tiene dos componentes; el primero es la *validez* del examen como instrumento de medida de aquello que se quiere realmente evaluar –el dominio de un campo de conocimiento, en nuestro caso-; el segundo componente hace referencia a la capacidad del examen para poner de manifiesto el dominio del alumno en

los aspectos más relevantes del aprendizaje: la capacidad de *comprensión y análisis* de la materia evaluada, la capacidad de asociación de conocimientos separados en una *síntesis* coherente y la *persistencia* de los conocimientos adquiridos. Un examen que cumpliera esos criterios –*validez*, discriminación de la *capacidad de análisis y síntesis* del estudiante y de la *persistencia* del conocimiento- sería un examen de *alta calidad*.

La *consistencia* representa la medida en la que cada parte del examen evalúa lo mismo que el examen en su totalidad. Es un concepto fundamental en las encuestas, pero tiene un valor más discutible en un examen que esté compuesto por múltiples preguntas –abiertas o cerradas-; en estos casos, depende de las características de la materia, o materias, que se evalúan. Como dijimos antes, en el capítulo del Marco Teórico, la consistencia interna de un examen tiene una relevancia relativa, y puede ser baja –o *no muy elevada*- sin que ello reste validez al examen. De todos modos, aun aceptando la afirmación anterior, que un tipo de examen pierda consistencia interna frente a otro **de la misma materia** indicaría una construcción defectuosa en el primero. Por eso, aunque de forma relativa, la consistencia interna de los exámenes conserva valor como criterio de calidad, de acuerdo con Mayer.

La *eficiencia* es la capacidad de conseguir los resultados deseados con el menor consumo de medios posible; un examen eficiente sería aquel que no exigiese excesivo tiempo en su preparación y corrección por parte del examinador y que no resultase innecesariamente arduo, dificultoso y agotador para el examinando. La utilización del sistema de test como método de examen ha buscado siempre aumentar la eficiencia del proceso evaluador sin disminuir la validez.

La *efectividad*, capacidad de conseguir el efecto deseado (concepto muy próximo al de eficacia) se puede incluir -en el campo de la evaluación- dentro del concepto de *validez*, pues solo es efectivo lo que es válido.

Siguiendo este esquema de Meyer, nosotros hemos estudiado el test basado en la confianza para evaluar su *calidad, consistencia y eficiencia*, entendidos en el sentido que hemos expuesto antes. Pero a esos criterios, necesarios para satisfacer los fines del evaluador, añadimos un cuarto elemento que nos parece el más relevante del proceso de aprendizaje: *los exámenes han de servir para orientar y estimular al alumno en su esfuerzo de aprendizaje*. Para poder afirmar que un sistema de evaluación es óptimo, no solo ha de tener *calidad, consistencia y eficiencia*, ha de tener una influencia detectable en el esfuerzo de aprendizaje del alumno, y esa influencia habrá de ser potenciadora del esfuerzo de estudio, del esfuerzo de comprensión y habrá de moverle a la adopción de las estrategias más adecuadas de aprendizaje.

Hemos sido conscientes de las fortalezas y debilidades del sistema de examen de tipo test, tal vez el más utilizado en nuestro medio, y hemos escogido el test basado en la confianza como un sistema de examen que podía mejorar las deficiencias de los test convencionales. Esas deficiencias, como expresamos en el capítulo segundo, eran el poco estímulo para el esfuerzo del estudiante y para un aprendizaje de calidad (Oh, 2004), las dudas en cuanto a su validez y fiabilidad, la poca discriminación de los alumnos brillantes (Hutchinson, 1982), la dificultad para elaborar buenos enunciados de preguntas (Burton, 2005; Cronbach, 1970; Haladyna y otros, 2002; Rød, 2010), el sesgo de género (Hassmen y otros 1994; Ben-Shakhar y otros 1991; Budescu y otros 1993), la tendencia a la «adivinación» en los alumnos (Haladyna, 2004; Oh, 2004) y la no valoración del conocimiento parcial (Morales, 2006a; Rippey, 1973). Consideramos que el sistema de test basado en el grado de confianza podría eliminar varias de esas

deficiencias de los test convencionales, concretamente la no valoración del conocimiento parcial y la tendencia a la «adivinación» en los alumnos; en el último caso no eliminándola sino ponderándola y haciéndola un componente más de la nota del examen. Por ello adoptamos, hace ya 6 años, el TBC como sistema principal de examen.

Al ser el TBC un sistema prácticamente nuevo en nuestro país, considerábamos que debíamos estudiar sus características; fundamentalmente su «calidad» -en el sentido que le da a ese término Mayer- y la adaptación de nuestros alumnos a ese procedimiento evaluador. Por ser el único profesor que utiliza el TBC –no conseguí que otros docentes lo utlizasen- he tenido que estudiarlo en los estudiantes a los que imparto clase, que son estudiantes de segundo y tercer curso de Ciencias de la Salud, en los grados de Enfermería y Fisioterapia, de una universidad privada, la Universidad Pontificia Comillas. Este carácter homogéneo de los participantes incluidos en el estudio limita la representatividad del mismo, pero creemos que sus resultados serán fácilmente extrapolables a otros grupos de estudiantes; en todo caso, es toda la población con la que contamos para hacer el estudio.

6.2. VALIDEZ

La primera característica del TBC que estudiamos fue la de su validez. Para ello utilizamos como variable-criterio los resultados de los alumnos en un examen de tipo «preguntas abiertas», y comparamos con esos resultados los obtenidos por el alumno con el examen de tipo TBC y con el test «convencional» por aciertos. El examen con mayor validez sería aquél que se aproximase en sus resultados más a los resultados obtenidos con el examen de preguntas abiertas. Para disminuir la varianza de los dos

sistemas de test, en este caso el test era único, corregido por los dos procedimientos (TBC y *por aciertos*), y fue realizado el mismo día y sobre la misma materia que el examen de preguntas abiertas.

Tal como se expuso en el Marco Conceptual (capítulo 2), en los estudios sobre la validez del TBC, escasos en los últimos años, realizados en otros países, varios autores han referido un incremento en la validez del TBC frente al test convencional. Así, Gardner (1969), Hamblton (1970), Kansup y Hakstian (1975) y Leclerq (1982), encontraron una correlación de los resultados obtenidos en las pruebas basadas en el nivel de confianza con los obtenidos con otros sistemas de evaluación -comprobación habitual de la *validez*- mayor que cuando se corrigen simplemente por el número de aciertos. Rippey, en cambio, en un trabajo publicado en 1972, no encontró incremento en la validez del test basado en la confianza frente a otros sistemas. **Nosotros, en los alumnos de Enfermería y Fisioterapia de nuestra universidad, hemos encontrado que el TBC es un sistema de examen con mayor validez que el test convencional por aciertos.** Tiene una correlación con el examen de preguntas abiertas significativamente mayor que la del test por aciertos. Este dato objetivo obtenido en 2008 y 2009, la mayor validez del TBC frente al *test por aciertos*, coincide con la opinión de los alumnos de 2013 y 2014, expresada en la encuesta, que consideran que el TBC es el sistema de examen que proporciona mejor información al profesor sobre el nivel de conocimientos del estudiante y al alumno le permite descubrir mejor sus lagunas formativas, características ambas que definen la validez de una evaluación. Así, la opinión de los estudiantes coincide, 5 años después, con los resultados del estudio de validez, que utilizó puntuaciones de los exámenes.

6.3. FIABILIDAD

El segundo punto que estudiamos en nuestro trabajo fue el de la consistencia interna, o fiabilidad, del TBC. Al igual que sucedía en el caso de la validez, en lo referente a la fiabilidad de los exámenes basados en la confianza había datos contradictorios en estudios anteriores. Para Hevner (1932), Ahlgren (1969), Anderson (1982) y Gardner–Medwin (2005), la consistencia interna (o fiabilidad) de los resultados aumentó en el test basado en la confianza. Por el contrario, Frary (1982) no encontró aumento de la fiabilidad frente al test convencional.

La fiabilidad de los exámenes de tipo test la medimos a partir del índice de consistencia interna α de Cronbach. Hay otros procedimientos para medir la fiabilidad, como la fórmula de Spearman-Brown (Morales, 2008, pp. 225-226), que resulta más engorrosa de calcular y es menos utilizada, o el sistema del *test-retest*, que es inaplicable en exámenes. Nosotros hemos utilizado el índice α de Cronbach. Pero ese índice, que es un valioso parámetro en una encuesta, tiene un significado mucho menos determinante cuando se trata de un examen. Un examen no necesita tener una alfa de Cronbach alta, ni siquiera mediana, para estar bien construido. Solo en algunas materias, como las matemáticas o el aprendizaje de idiomas, la consistencia interna tiene un auténtico valor como medida de la buena construcción del examen. Es decir, en materias en donde hay un único campo de conocimientos que se va desarrollando paulatinamente y en el que todos los conocimientos están vinculados, donde hay una única dimensión a medir, el análisis de la consistencia interna tiene auténtico significado y valor. En materias que, como ocurre en la nuestra, consisten en múltiples campos de conocimiento que pueden ser independientes unos de otros, la consistencia interna puede no ser muy alta, incluso puede ser un tanto baja, sin que ello afecte a la

validez. En definitiva, la medida de la consistencia interna nos dice hasta qué punto todos los ítems miden lo mismo, de manera que la respuesta en un ítem aislado tiene un valor predictivo del resultado final. En un examen, eso se traduce en el hecho de que tener respondida correctamente –o incorrectamente en el caso opuesto- una pregunta nos aporta una predicción (limitada, pero real) del resultado final. En materias en las que todos los conocimientos suelen estar estrechamente vinculados, como en los casos que mencionamos antes de Matemáticas o un idioma, cada pregunta nos proporciona una predicción veraz del resultado final; por poner un ejemplo del campo de las matemáticas: si el estudiante no resuelve un problema elemental seguramente tampoco podrá resolver los problemas complejos, si resuelve un problema complejo seguramente habrá resuelto con facilidad los elementales; de esta manera cada ítem, con su peculiar dificultad, se convierte en un predictor del resultado global. Pero en una materia como la Patología, que constituye la esencia de nuestras asignaturas, un estudiante puede dominar exhaustivamente un campo –enfermedades respiratorias, por ejemplo- y desconocer totalmente otro –Neurología, por caso- y en otro estudiante puede suceder exactamente lo contrario; y así tener ambos la misma puntuación y no coincidir en ningún ítem acertado; la varianza total del examen puede ser alta, pero la covarianza de los ítems es baja, lo que, de acuerdo con las fórmulas del coeficiente α de Cronbach, el más utilizado para el cálculo de la consistencia interna, implica una consistencia interna reducida (Morales, 2008, pg. 193-212); pero esa baja consistencia interna no afecta a su validez como instrumento de medida de unos conocimientos determinados. No es una disminución de la validez del examen lo que señala esa alfa reducida, simplemente indica una escasa precisión de la medida (Morales, 2008), lo que se puede compensar con un número elevado de ítems. En las materias que están formadas por una suma de conocimientos parciales poco relacionados, la consistencia

interna de los exámenes tiende a ser baja, pero eso no significa necesariamente que estén mal diseñados.

También aquí hicimos la comparación con el test por aciertos y nuestra hipótesis era la de que el TBC mantenía, al menos, la consistencia interna del test por aciertos. La razón de considerar suficiente como cualidad del TBC el mantener una consistencia interna similar a la de los test convencionales radicaba en que el TBC no valora solamente un rasgo del alumno –conocimiento-, como sí hace el test convencional por aciertos, sino que valora dos dimensiones, conocimiento y confianza. Aunque los conocimientos y la confianza sean dos variables relacionadas, son variables diferentes y podrían disminuir el índice α de Cronbach del TBC frente al convencional, que es unidimensional. Nuestra hipótesis es que la consistencia interna no disminuye en el TBC.

Nuestros resultados muestran que el sistema del TBC no solo no disminuye la consistencia interna del examen frente al test convencional, sino que la aumenta en todos los casos evaluados menos en uno, en el que el índice α de Cronbach es igual en ambos tipos de examen. La diferencia de alfas es significativa en uno de los grupos estudiados, el más numeroso con diferencia –lo que no es casual-, pero el hecho de que en los demás exámenes sea más elevada –con la excepción del caso en que ambos son iguales- indica una tendencia en el TBC a elevar la fiabilidad frente al test convencional. La fiabilidad no es una característica intrínseca del instrumento –el tipo de test en este caso- sino de los resultados obtenidos con él (Morales, 2008, pg. 190-191); por ello no se pueden sumar los datos de los cinco exámenes en los que se ha estudiado; lo correcto es presentarlos separadamente, como nosotros hemos hecho, con el inconveniente de que en cuatro de los cinco grupos el número de participantes, entre

28 y 44, era demasiado reducido como para esperar significación estadística de la diferencia.

6.4. ACEPTACIÓN DEL TBC POR EL ESTUDIANTE

Una cualidad importante que ha de tener un sistema de examen es la de no provocar un rechazo excesivo en el estudiante. La buena aceptación del procedimiento evaluador hará que el alumno se implique más en él y que le pueda servir como estímulo positivo para el aprendizaje. Un examen que genere rechazo previsiblemente aumentará el estrés del alumno al enfrentarse a él. Nosotros encontramos entre nuestros estudiantes una aceptación mayor del test convencional, basado en el número de aciertos, que del test basado en la confianza. Pero aunque su aceptación sea inferior a la test por aciertos, al que los estudiantes están más habituados, el TBC no provoca un rechazo en la mayoría de nuestros alumnos –el 54 % de ellos le dan puntuación positiva-. La mayoría de los participantes prefieren el TBC al examen de preguntas abiertas.

Aunque no tanto como el test convencional, es TBC es un examen bien aceptado por los participantes en el estudio. En un estudio publicado por Swartz (2006), el autor encontró similares niveles de aceptación entre el test basado en número de aciertos y el test basado en la confianza.

El TBC es un sistema que, pese a que añade un grado de complejidad suplementaria al test por aciertos, es fácilmente comprendido en sus objetivos y su mecánica, por el estudiante. Hemos podido comprobar que el alumno mejora su habilidad para usarlo y es más certero en su elección del grado de seguridad a medida

que se habitúa a este tipo de examen. El sistema de puntuación que hemos utilizado, con tres categorías de confianza, ligeramente modificado de Gardner–Medwin (1995), es comprendido intuitivamente con gran facilidad por el estudiante. Otros sistemas de test basados en la confianza, como el de Hassmen y Hunt (1994) o el de Sainsbury y Walker (2008), tienen demasiados niveles de seguridad, cinco, y puntuaciones incoherentes o, en el primer caso, fraccionales, difíciles de recordar por el examinando. El sistema de Gardner-Medwin, con sus tres niveles de confianza, puntuaciones enteras sencillas y absoluta coherencia, es simple, comprensible e intuitivo.

6.5. EXÁMENES Y ESTRÉS DEL ESTUDIANTE

Coherentemente con el resultado obtenido en la pregunta que valoraba las preferencias del alumno por cada examen, la encuesta sobre el estrés que generan en el estudiante los exámenes muestra que el test por aciertos es el que genera menos estrés, el examen de preguntas abiertas es el más estresante y el TBC se encuentra en un punto medio entre los dos, con diferencias significativas en los tres casos.

Se ha comprobado que la ansiedad del alumno ante el examen es mayor cuando se dan las siguientes tres circunstancias: que el examen sea más importante, que el alumno tenga la expectativa de que se trate de un examen más difícil y que las condiciones del examen se perciban como poco familiares (Urosa, 1995). Es evidente que al enfrentarse a un test basado en la confianza el estudiante se encuentra en la segunda y tercera de esas situaciones estresantes; percibe que es un tipo de examen si no más difícil, al menos más exigente, y que es un examen diferente a los que está acostumbrado. Por ello es razonable que los estudiantes den mayor puntuación como estresante al TBC que al

test convencional por aciertos. Para atenuar la ansiedad frente al TBC, nosotros intentamos que el estudiante se familiarice con él, entrenándolo con pequeños exámenes de TBC previos durante un trimestre.

6.6. INFLUENCIA DEL TBC EN LOS HÁBITOS DE ESTUDIO DEL ALUMNO

Una cualidad que debe tener todo sistema de evaluación es la de influir en el estudio del alumno en el sentido de favorecer su aprendizaje. En este punto, un motivo de crítica a los exámenes de tipo test es el hecho de que puedan inducir al estudiante a realizar un menor esfuerzo de aprendizaje. Esto sucede porque al ser las respuestas cerradas, y estar ya redactadas, puede no exigir del estudiante un dominio tan importante de la materia como el examen de preguntas abiertas (Pintrich, 2002), por la posibilidad de acierto por azar (Rippey, 1978; Downing, 1992), por las pistas que facilitan la identificación de la respuesta correcta (Wolf, 1995, Haladyna y otros, 2002) o incluso porque pueda inducir un enfoque superficial de aprendizaje (Gardner-Medwin, 2006). Estas son críticas que se hacen reiteradamente a los sistemas de exámenes de tipo test.

Nuestros alumnos han opinado que el tipo de examen que esperan tener influye de forma importante en su forma de estudiar. Los participantes en el estudio expresaron que el TBC les obligaba a dedicar más tiempo al estudio, a memorizar más y a hacer un mayor esfuerzo de comprensión que el test por aciertos. El examen de preguntas abiertas es, en su opinión, el que les exige dedicar más tiempo al estudio, pero el TBC lo supera en cuanto a la capacidad de estimular el esfuerzo de comprensión de lo que se aprende. Por consiguiente, el aplicar el sistema de corrección basado en la confianza

permite corregir, a juicio de los alumnos, esa tendencia a un menor esfuerzo de aprendizaje que los autores han señalado en el caso de los exámenes de tipo test. Y puede estimular un aprendizaje basado en la comprensión más que los otros sistemas.

La opinión de que el TBC es el tipo de examen –de los tres encuestados- que fuerza al estudiante a intentar comprender mejor lo que aprende tiene gran relevancia, pues esa cualidad, fomentar la comprensión, es la que en la encuesta más se correlaciona con las opiniones favorables sobre la validez y la eficacia docente de los exámenes. Los alumnos asocian la calidad de un sistema de examen con su capacidad para estimular el esfuerzo de comprensión.

6.7. OPINIONES DE LOS ALUMNOS SOBRE LA VALIDEZ Y LA EFICACIA DOCENTE DEL TBC

Aunque el TBC no sea preferido por los estudiantes al test convencional por aciertos, en nuestra encuesta ellos muestran una opinión clara en el sentido de que es un sistema de examen mejor que los otros dos encuestados. En el agrupamiento de tres ítems que hemos llamado «*validez autopercebida*» del examen, que valora la opinión del estudiante sobre la *utilidad* del procedimiento de examen, expresada como la capacidad de cada tipo de examen para informar al profesor y al propio alumno sobre su nivel de conocimiento de la materia y hacerle aprender más, el TBC obtiene una puntuación significativamente superior al test por aciertos y al examen de preguntas abiertas. Por lo tanto, los alumnos expresan la opinión del que el TBC es un sistema de examen más útil –de superior «*validez autopercebida*»- que los otros dos.

En nuestra encuesta hemos creado un factor para estudiar la *calidad percibida por el alumno* de los distintos sistemas de examen; lo llamamos *eficacia docente*. El constructo que llamamos *eficacia docente* es una escala de 5 ítems, los tres de la *validez autopercebida* y los dos ítems que preguntan por el esfuerzo de estudio y de comprensión que cada examen exige. Es una pequeña encuesta que muestra una excelente consistencia interna, con valores de α que han llegado a 0.89. Creemos que este factor expresa adecuadamente *la calidad* de un sistema de examen, tal como la percibe el alumno. Al someter los tres tipos de examen a esta encuesta, encontramos que el TBC supera significativamente en eficacia docente al test convencional y al examen de preguntas abiertas. El alumno, aunque manifieste que le gusta más el test por aciertos que el TBC, sin embargo otorga la mayor puntuación al TBC cuando se le pregunta por su *calidad* como sistema de evaluación. Con independencia de su preferencia personal por un tipo de examen, los estudiantes encuestados, los cuales han tenido cuatro meses de experiencia previa con el test basado en la confianza, tienen una opinión excelente de él como sistema de evaluación, por encima de los demás exámenes.

6.8. TBC Y GÉNERO

Una cualidad necesaria en un sistema de evaluación es la de que no tenga sesgos en sus resultados en relación con las características personales del alumno. En nuestro medio, donde otros sesgos, como la etnia o raza de las personas, no suelen valorarse, el sesgo de género, su existencia o inexistencia, es un rasgo importante a estudiar.

Estudios previos habían mostrado una mayor tendencia de los varones a asumir riesgos en los exámenes (Zimmerman y otros, 1990; Ben-Shakhar y otros, 1991; Budescu y otros, 1993; Hassmen y otros, 1994; Urosa 1995, y Morony y otros, 2013). Pero otros estudios, basados en la comparación entre confianza y aciertos en los exámenes, han mostrado que, aunque los varones tienden a ser algo más arriesgados, o «sobre-confiados» que las mujeres, la diferencia es muy pequeña y no significativa estadísticamente (Pallier, Wilkinson, Danthiir, y otros, 2002; Pallier, 2003; Stankov y otros, 2012).

En un estudio reciente sobre la autoconfianza durante examen, Morony y otros (2013), utilizando una metodología similar basada en el «sesgo de puntuación» - diferencia entre la puntuación real obtenida en las preguntas de un examen y la puntuación esperada por el alumno (Moore y Haly, 2008)-, encontraron que los estudiantes varones tenían más confianza que las estudiantes mujeres en el campo de las matemáticas; la diferencia era considerable (un 19.7% en «Puntuación en confianza» a favor de los hombres), significativa y relevante, al haberse hecho el estudio en una muestra de más de 7000 sujetos. Otros estudios, más limitados en número de sujetos, encontraron también hallazgos similares a los del anterior (Chen, 2003; Ramdass y otros, 2008).

Nosotros, en nuestro trabajo, tanto en el estudio preliminar realizado en 2008-2009 como en el posterior, de 2013-2014, no hemos encontrado diferencias significativas entre ambos géneros en su elección del nivel de confianza. Utilizando el parámetro PC-aciertos, que pone en relación el nivel de confianza con el nivel de conocimientos, constatamos que los varones y las mujeres no tenían diferencias significativas en su tendencia a arriesgarse. La contradicción de nuestros resultados con

los del estudio de confianza en el examen PISA merece comentarse. El estudio de confianza de Morony se realizó en estudiantes de secundaria durante los exámenes PISA de matemáticas. Los estudiantes eran de 10 nacionalidades diferentes, 5 de Europa y 5 de Asia. En la menor confianza de las mujeres en ese estudio ha podido influir el que fuera un examen de matemáticas, donde los resultados académicos de las estudiantes femeninas fueron significativamente peores que los de los varones, así como que la mitad de los participantes pertenecieran a países del Extremo Oriente, donde los condicionamientos sociales no favorecen en la mujer la actitud de confianza, frente a lo que ocurre con los varones (Morony y otros, 2013). Nosotros no encontramos sesgo de género en nuestros estudios en lo referente al nivel de confianza. El número de participantes en nuestros estudios, hechos con cinco años de intervalo en 2009 y 2014, no fue muy elevado; sumando ambos, en total contamos con 227 participantes, de ellos 158 mujeres y 69 hombres -88 en 2009 y 139 en 2014-. Pero en ambos casos no hubo diferencia significativa entre el nivel de confianza (parámetro PC-aciertos) de hombres y mujeres, con la característica de que en los dos estudios el nivel de confianza de las mujeres fue ligeramente más elevado –aunque no significativamente- que el de los hombres. También da consistencia a nuestro trabajo el gran paralelismo de los resultados, que han sido prácticamente iguales en hombres y mujeres al comparar el estudio de 2008-2009 y el de 2014 –diferencias de solo un 6 % en mujeres y nulas en varones-. El hecho de que hayan pasado 5 años entre uno y otro, aumenta la sensación de validez de nuestros datos. Una característica de nuestros grupos de estudiantes, que supone una dificultad para hacer estudios comparativos según el género, es el número relativamente reducido de varones que cursan nuestros grados de Ciencias de la Salud, especialmente en Enfermería, donde el 85% de los participantes son mujeres.

En nuestro estudio, las chicas se arriesgaron en la misma medida que los chicos. Hay que destacar que nuestros estudiantes pertenecen al área de Ciencias de la Salud, campo donde la presencia femenina es ampliamente mayoritaria, que tenían de media 4 años más que las que estaban haciendo el examen PISA –factor no desdeñable cuando se trata de valorar seguridad o inseguridad- y que la elección del grado de seguridad en nuestro estudio no es una encuesta paralela al examen, sino que forma parte del propio examen y condiciona fuertemente la puntuación final, por lo que entran en juego factores estratégicos que en el estudio de Morony, una encuesta sin repercusión para el participante, no entraban. Por ello no ponemos en duda nuestros resultados, contradictorios con el estudio antes mencionado, a pesar del menor número de sujetos participantes. Si interesa profundizar en el tema, será necesario repetir el estudio con otro tipo de estudiantes, de otras titulaciones y otras edades.

Nos parece que sería interesante estudiar más a fondo esta cuestión, la ausencia o presencia de sesgo asociada al género en la tendencia a sumir riesgos en exámenes, en posteriores estudios. El sistema de examen basado en la confianza, que nos da todos los datos que necesitamos para conocer la tendencia al riesgo de cada alumno, sería un procedimiento excelente para realizar esos estudios.

En cuanto a la habilidad en la utilización del TBC, la medida en que beneficia o perjudica a un alumno este sistema de examen, nuestro estudio no mostró diferencias significativas entre mujeres y hombres. El parámetro que utilizamos para medir esta habilidad es el *TBC-aciertos*. Un *TBC-aciertos* positivo nos indica que el estudiante obtuvo mejor nota con el TBC de la que hubiese obtenido si en el examen solo hubiera contado el número de preguntas acertadas o falladas; un *TBC-aciertos* negativo indica lo contrario. En nuestro estudio no hubo diferencias significativas en el *TBC-aciertos*

entre estudiantes mujeres y varones. El TBC no tiene sesgo de género, no favorece ni perjudica a ninguno de los dos sexos.

En la encuesta sí hemos encontrado una diferencia significativa según el género de los participantes; las mujeres referían un mayor nivel de estrés al afrontar los exámenes que los varones. Este hecho sucedía con los tres tipos de examen encuestados, -TBC, test por aciertos y preguntas abiertas- y la diferencia era significativa ($p < 0.001$) y de magnitud considerable –tamaño del efecto de 0.60 a 0.90-. Las estudiantes reconocen un mayor nivel de estrés ante los exámenes que los estudiantes del género masculino. En qué medida ese dato responde a un nivel de estrés real o expresa solo una tendencia mayor en la mujer que en el varón a reconocer sus padecimientos es algo que en el presente trabajo no podemos dilucidar, pero otros estudios han corroborado también la mayor percepción de estrés ante un examen en las mujeres (Morony y otros, 2013).

6.9. TBC Y ENFOQUES DE APRENDIZAJE

De acuerdo con la definición de Torre (2007), los enfoques de aprendizaje son la representación de la forma en que los estudiantes afrontan sus actividades académicas, tanto en sus motivaciones como en sus estrategias. El *enfoque profundo* se esfuerza en buscar enlaces lógicos, relaciones entre los conceptos y significados; busca profundizar en el conocimiento, *apropiárselo*, más allá de su utilidad para probar exámenes. El enfoque superficial de aprendizaje, en cambio, está basado en la repetición de los conceptos hasta que estos quedan memorizados, buscando solo responder a las preguntas de los exámenes sin pretender alcanzar significados globales; el alumno con un enfoque superficial de aprendizaje no pretende apropiarse personalmente de

conocimientos, solo reproducirlos (Kember y Harper, 1987; Pérez y otros, 2011; Rosario, 2005, 2007; Torre, 2005). El aprendizaje de calidad exige enfoques profundos por parte del estudiante.

Desde que se empezaron a utilizar por los investigadores los conceptos de enfoques de aprendizaje, se han realizado numerosos estudios para conocer su relación con los resultados académicos del estudiante. Como señalamos en el capítulo 2, algunos autores (Ekins 1991; Gargallo, 2006; Papinczak, 2009; Svenson, 1977; Torre, 2005) encontraron correlaciones significativas entre los enfoques profundos o profundo-estratégicos de aprendizaje y unos mejores resultados académicos, aunque la intensidad de las correlaciones no fue muy grande. Otros autores, en cambio, encuentran que los enfoques profundos como predictores de los resultados en exámenes y presentaciones orales resultan pobres (Recio, 2005; Swanberg, 2010; Vrugt, 2008). Parece claro que la correlación entre enfoques de aprendizaje y resultados académicos, especialmente con los resultados de los exámenes, existe, pero es pequeña. Una circunstancia que podría explicar la escasa capacidad predictiva de las notas en exámenes que los enfoques profundos presentan frente a los superficiales podría ser el hecho de que los estudiantes con enfoques superficiales tienen su foco de interés académico centrado exclusivamente en los propios exámenes, concentrando su esfuerzo en poder responder a las preguntas de estos, y conseguir las mejores puntuaciones posibles en relación a sus conocimientos; los alumnos con enfoques profundos, en cambio, tienen un interés académico que va más allá de los propios exámenes y se centran en objetivos posteriores y más trascendentes. De manera que podría darse la situación de que nuestros sistemas de examen realmente incentivasen el aprendizaje superficial y desincentivasen el profundo. Cuando elaboramos evaluaciones que exigen del alumno reproducción del conocimiento, pero no un alto nivel de comprensión y asociación de conceptos,

podemos estar estimulando en él los enfoques superficiales. Por este motivo, Tian (2007) afirma que un buen sistema de evaluación debería reconocer y premiar los enfoques profundos. Shoendorfer y Emmett, en 2012, encontraron que la utilización de un sistema de examen basado en la confianza en lugar del test tradicional mejoraba los resultados académicos de los estudiantes. En este sentido, algunos autores han sugerido que el añadir una segunda dimensión al conocimiento –como es el caso de la confianza- en las evaluaciones puede promover un nivel más profundo de reflexión y autorregulación en el alumno y, por lo tanto, estimular enfoques profundos y mejorar los resultados del aprendizaje (Adams y Ewen, 2009; Stankov y otros, 2012).

Nosotros hemos estudiado la relación entre los enfoques de aprendizaje, tanto profundos como superficiales, con los resultados académicos en nuestras asignaturas, y hemos estudiado especialmente la relación entre el TBC y los enfoques profundos y superficiales. Consideramos que si el TBC premiara con mayor puntuación a los alumnos que declaraban enfoques profundos de aprendizaje podría servir de estímulo para la adopción de éstos por el estudiante, evitando el efecto perverso de estímulo al aprendizaje superficial que los exámenes centrados en la reproducción del conocimiento pueden tener. En consecuencia, si probásemos esta relación -enfoques profundos/buenos resultados en el TBC- ello afirmaría la calidad del TBC como sistema de evaluación por encima de los test convencionales.

Para estudiar la relación TBC-enfoques de aprendizaje dispusimos de tres instrumentos:

- La escala R-SPQ-2F, de Biggs, que valora enfoques profundos y superficiales en el estudiante -con los subfactores *motivo* y *estrategia*-.

- La encuesta en donde el alumno valoraba comparativamente tres sistemas de evaluación –TBC, test por aciertos y preguntas abiertas-.
- Las puntuaciones en el examen de tipo test que el alumno había realizado unos dos meses antes. De ese examen disponíamos de tres parámetros, la puntuación según el sistema TBC, la puntuación según número de aciertos y la diferencia TBC-aciertos; esta última nos indica si el alumno sale beneficiado o perjudicado por el sistema basado en la confianza.

Hemos encontrado en nuestro trabajo una asociación entre el TBC y el aprendizaje profundo desde dos puntos de vista:

- Los alumnos que tienen mejor opinión del examen basado en la confianza – que le otorgan mayor *eficacia docente* y tienen preferencia por él- son los que obtienen mayor puntuación en enfoques profundos de aprendizaje, mientras que los que tienen peor concepto del TBC son los que tienen enfoques de aprendizaje superficiales.
- Al comparar los resultados de las evaluaciones con los enfoques del alumno se comprueba que los alumnos con enfoques profundos de aprendizaje (especialmente estrategias profundas) se desenvuelven mejor con el TBC y son premiados por ello con mayor puntuación de la que obtendrían si hubieran sido examinados mediante un test convencional *por aciertos*.

Encontramos que la alta puntuación en el test de Biggs de los enfoques profundos se correlacionaba significativamente, y habitualmente con un nivel alto de r , con tener una opinión positiva sobre el TBC. Los parámetros que expresan una opinión positiva de los exámenes son la preferencia del alumno por ese examen –ítems 2,3 y 4-, el

estímulo que el examen presta al esfuerzo de comprensión de conceptos –ítems 8, 9 y 10- , y los constructos que llamamos *validez autopercebida* y *eficacia docente* del examen. En todos esos parámetros hubo una correlación positiva significativa entre otorgar una alta puntuación al TBC y tener enfoques profundos de aprendizaje, y una correlación negativa significativa entre los enfoques superficiales y otorgar una alta puntuación al TBC. Con los otros dos tipos de examen no se observaron en la mayoría de los casos correlaciones significativas con los enfoques de aprendizaje y en los pocos casos en que aparecieron, fueron de cuantía mucho más baja que las que hubo con el TBC (tabla 4,58 y 4.63). Por consiguiente, los alumnos con enfoques profundos, motivo y estrategia, tienden a tener una opinión positiva del TBC como sistema de evaluación: tienen mayor preferencia por él, le reconocen más validez y eficacia docente.

En nuestro trabajo demostramos que los alumnos con alta puntuación en *enfoques profundos* de aprendizaje, especialmente en *estrategias profundas*, tienen una habilidad mayor en la utilización del TBC que los que no tienen esa alta puntuación. Este hallazgo se puso de manifiesto en la correlación positiva significativa entre el parámetro *TBC-aciertos* (nota obtenida en el TBC menos nota en la corrección por *número de aciertos*) y la alta puntuación en enfoques profundos de aprendizaje y, dentro de ellos, en estrategias profundas. Repetimos la magnitud de esa correlación entre estrategias profundas y habilidad en la utilización del TBC: r de **0.23** ($p < 0.001$). La correlación entre los enfoques profundos (motivo + estrategia) y el TBC-aciertos también fue positiva y significativa, aunque un poco menor (r de 0.20, $p < 0.01$). Esto significa que los alumnos que utilizaron estrategias profundas en la preparación del examen tuvieron con el TBC de promedio una puntuación más elevada que la que hubieran tenido si el examen hubiese sido corregido solo mediante el número de aciertos. **El TBC reconoce y premia con un aumento de puntuación los enfoques profundos de aprendizaje**

frente al test convencional por aciertos. Esta correlación se estableció entre un valor en una escala de opinión,—el test de Biggs, aceptado como instrumento válido y fiable para medir enfoques de aprendizaje— y la puntuación del examen, examen que los alumnos habían hecho dos meses antes de la encuesta; correlación entre un dato objetivo y una medida en una escala de actitudes. Teniendo en cuenta la cantidad de variables que entran en juego a la hora de responder a una encuesta, y más dado el tiempo transcurrido entre examen y test de Biggs, que se haya podido encontrar esa correlación nos parece muy relevante. Creemos que los datos obtenidos — r de 0.20 con enfoques profundos, r de 0,23 con estrategias profundas— son la «punta del iceberg» de la auténtica correlación entre TBC y enfoques de aprendizaje, que probablemente sea mayor en la realidad.

Esa correlación -test basado en la confianza con enfoques profundos de aprendizaje— es explicable lógicamente conociendo las características del TBC, que se basa en la evaluación conjunta del conocimiento y la seguridad. En un examen, un alumno está completamente seguro de la respuesta elegida cuando es capaz de razonarla, de dar argumentos científicos o lógicos a favor de ella, y mucho más si también puede razonar por qué las otras respuestas son falsas; de manera que la puntuación en este tipo de exámenes tiene, de alguna forma, relación con el aprendizaje profundo, que hace hincapié en significados y relaciones de los conocimientos, más que en la memorización de datos.

Por consiguiente, si el TBC premia con mayor puntuación los enfoques profundos de aprendizaje (en mayor medida que el test convencional), cabe pensar que el uso del test basado en la confianza pueda estimular al alumno a adoptar estrategias de aprendizaje profundo —buscar significados, enlaces lógicos y relaciones entre conceptos,

apropiándose del conocimiento- pues esas estrategias le permitirán obtener mejores resultados. Al estimular esas estrategias, el TBC aporta un aumento de la *calidad evaluativa*, en el sentido que dimos antes al término de *calidad*, basándonos en las definiciones de Mayer (2002).

Como vimos más arriba, numerosos autores han encontrado una correlación significativa entre los enfoques profundos de aprendizaje y los resultados académicos – correlación generalmente pequeña- (Ekins 1991; Gargallo, 2006; Papinczak, 2009; Svenson, 1977); nosotros también la hemos encontrado, pero lo que aportamos es la constatación de que al introducir la valoración de la confianza en el test, esa correlación aumenta significativamente.

6.10. TBC Y AUTORREGULACIÓN

En la encuesta introdujimos una escala de autorregulación de 5 ítems; mediante esta escala, que ha tenido una buena consistencia interna, hemos podido demostrar la correlación entre autorregulación y la confianza académica, expresada en el TBC.

La autorregulación es un elemento fundamental en el proceso de aprendizaje del estudiante. Como vimos en el segundo capítulo, Pintrich (1995) e Isaacson y otros (2006) destacan que la autorregulación no es solo la aplicación de unas determinadas estrategias de aprendizaje, sino que es un proceso complejo que exige autorreflexión y autoconocimiento, así como conocimiento de las tareas académicas, motivación y autocontrol activo del comportamiento. En el proceso de autorregulación, la autorreflexión y la automonitorización son elementos críticos a lo largo de todo el

proceso; los alumnos conscientes de su nivel de maestría pueden ajustar sus estrategias y prácticas de estudio, pueden autorregularse eficazmente (Kelemen, 2000). El aprendizaje no se puede autorregular si no es a través de un proceso de autoconocimiento (Winne, 1996) y la confianza es la mejor medida del autoconocimiento (Sharma y otros, 2011); por ello la introducción de la confianza como componente de la nota de una evaluación puede hacer que el resultado académico refleje mejor la autorregulación del alumno. Y que un examen reconozca y califique una determinada actividad supone que la potencie. Si, como dice Pedro Morales, en las actividades de aprendizaje lo que no se *evalúa* de *devalúa*, entonces lo que se *evalúa* tiende a *revalorizarse*. Hemos obtenido en el estudio una correlación significativa entre TBC-aciertos y puntuación alta en la escala de autorregulación (r de 0.15, $p < 0.05$). La correlación estadísticamente significativa entre la alta puntuación en la escala de autorregulación y el parámetro TBC-aciertos significa que el TBC reconoce y premia, con mayor puntuación que el test convencional, las estrategias de autorregulación, al igual que hacía con las estrategias de aprendizaje profundo. Al evaluarlas y premiarlas puede incentivarlas en los estudiantes.

Los alumnos con mayor puntuación en autorregulación también son los que otorgan mayor puntuación al TBC como buen sistema de examen (mayor puntuación en *eficacia docente*, en *validez autopercebida* y en *estímulo a la comprensión* de la materia con el TBC), correlación que no se da con los otros exámenes. Estos hallazgos refuerzan la relación entre las actividades de autorregulación del aprendizaje y la autoconfianza del estudiante.

6.11. MEDICIÓN DE LA CONFIANZA ACADÉMICA CON EL TBC

La confianza, como dijimos en el capítulo 2, está siendo objeto de gran interés en los últimos años como un componente esencial del proceso de aprendizaje con capacidad predictora de los resultados que obtienen los estudiantes (Kleitman y Stankov, 2001; Morony y otros, 2013; Nix y otros, 2011; Stankov y otros, 1997; Stankov y otros, 2012;). Para el estudio de la confianza se utilizan encuestas de confianza académica, como la escala de confianza ABC (Sander y Sanders, 2003). En nuestros exámenes TBC nosotros medimos la confianza de los alumnos utilizando para ello su opción de confianza, en la escala de 10 (plena confianza), 5 (confianza intermedia) y 0 (ausencia de confianza). Esta puntuación se asigna a cada pregunta del test y la media de todas ellas, nos da la puntuación en confianza del examen del estudiante. La confianza del alumno es la suma de su confianza en cada una de las preguntas del examen; la confianza es un constructo aditivo (Bong y Skaavik, 2003). Nuestro sistema de calcular la confianza en el examen es marcadamente similar al utilizado por Morony y otros en su estudio de la confianza en el examen de matemáticas PISA, publicado en 2013; también ellos valoran la confianza en una escala de 0 a 10 en cada pregunta del examen y construyen la confianza media con la suma de las puntuaciones en confianza de cada pregunta. La diferencia entre nuestra medida de confianza y la suya proviene del hecho de que el instrumento de Morony es una mera encuesta, sin repercusión en los resultados académicos, mientras que nuestra medida de la confianza del alumno en cada pregunta es un componente de la puntuación que el alumno obtiene. Esa trascendencia del nivel de confianza en la calificación hace que entren en juego factores estratégicos por parte del alumno que quiere optimizar su calificación, que no existen en una simple encuesta. Pero tanto la escala de Morony como nuestro procedimiento de calcular la confianza en el TBC resultaron buenos

predictores de los resultados académicos; en el caso de Morony la puntuación en confianza se correlacionaba con los resultados del examen con una r de Pearson de 0.60 (Morony y otros, 2013), mientras que la correlación de la confianza con las puntuaciones del examen fue de 0.70 y 0.68 en las dos mediciones hechas en nuestro caso (2009 y 2014).

Tanto Morony como nosotros relacionamos la puntuación en confianza con la nota obtenida en el examen y así obtenemos un índice que nos indica la tendencia al riesgo o a la prudencia del estudiante ante el examen. El índice se basa en el trabajo de Moore y Haly (2008), que comparan la nota esperada por el estudiante con la nota real. Morony, al igual que nosotros, no compara con el resultado del examen una estimación global hecha por el alumno de su nota, como hace Moore, sino que en cada pregunta o problema del examen –problemas y preguntas cortas en su caso, en el nuestro preguntas de test- el alumno tiene que expresar su confianza en haberla respondido bien, siendo la suma de estas confianzas parciales la que da la puntuación en confianza global. Ellos suelen llamar «sobreconfianza», o también «sesgo de confianza», a este índice; nosotros, en el presente estudio, no le hemos adjudicado un nombre específico, sino que la hemos denominado simplemente índice *PC-aciertos*. Cuando sale un valor del *PC-aciertos* positivo, superior a 0, el estudiante tiene una confianza superior a sus resultados, cuando sale negativo sucede todo lo contrario.

La discrepancia entre *el conocimiento y la confianza* de un alumno nos sugiere la cuestión, *¿por qué existe esa discrepancia?* Aunque considerando el grupo de alumnos como un todo esa discrepancia prácticamente es nula (nota media en el examen 6.16, puntuación media en confianza 6.11, diferencia no significativa), muchos alumnos tienen grandes diferencias entre conocimientos y confianza. *¿Qué es lo que lleva a un*

alumno a mostrarse demasiado inseguro o excesivamente seguro en relación a sus conocimientos?

Hay tres factores que pueden hacer que exista la discrepancia conocimientos–seguridad. El primero de ellos, señalado ya por Gardner–Medwin (2006), puede ser la falta de sinceridad, es decir, que el alumno no muestre su auténtica confianza al optar por su grado de seguridad, sino que la exagere o la disminuya por motivos estratégicos. El segundo factor sería el auténtico desconocimiento del nivel de sus propios conocimientos, el déficit en autorreflexión y metacognición (Efklides, 2011). El tercer factor vendría dado por la psicología del sujeto y tendría que ver con su tendencia personal hacia la prudencia o hacia la asunción de riesgos, la tendencia a arriesgarse como un rasgo de su personalidad. Para Stankov la confianza se encuentra en una “tierra de nadie” entre las habilidades cognitivas y los aspectos de la personalidad del estudiante (Stankov y otros, 2012).

En nuestra opinión, el estudio sugiere que la temeridad de los estudiantes con menor nota y mayor confianza se debe probablemente en la mayoría de casos a la falta de sinceridad (o cálculo estratégico para intentar conseguir la mejor nota posible con escasos conocimientos, riesgo asumido por *desesperación*), al igual que pensaba Gardner-Medwin (2006a); este hecho, la asociación entre un bajo nivel de conocimientos y la tendencia a aumentar el riesgo relativo, ha sido detectado también en otros estudios sobre auto–confianza (Kasperski y Ketziu, 2013); los autores llegan a la misma conclusión que nosotros, los alumnos con menos conocimientos han de arriesgarse más si quieren aprobar el examen.

La discrepancia entre confianza y conocimientos en sentido contrario al anterior, por excesiva prudencia –poca *confianza relativa*–, que es mayor entre los alumnos bien pero no sobradamente preparados, se debería tanto a un cálculo estratégico (asegurar el aprobado sin correr riesgos) como a un auténtico defecto de su metacognición, un fallo en la automonitorización del aprendizaje. Estos alumnos reducen su nivel de confianza una vez que comprueban que con las respuestas aseguradas han garantizado el aprobado; arriesgarse más podría beneficiarles en la nota, pero con mala suerte podría perjudicarles y poner en peligro sus objetivos mínimos. Aunque esto no deja de ser un cálculo estratégico, también expresa algo más, un déficit en su metacognición, pues nadie renunciaría voluntariamente a obtener una nota más alta, lo que conseguiría escogiendo opciones mayores de confianza, si no dudara realmente de la firmeza de sus conocimientos. De esta manera, un índice PC–aciertos positivo y alto en un alumno con baja nota parecería indicar falta de sinceridad en el examinando, mientras que un índice PC–aciertos muy negativo en un alumno con buena nota indicaría, además, un déficit de metacognición, un desconocimiento de la medida y los límites del propio conocimiento.

El hecho, constatado reiteradamente desde que empezamos a utilizar este sistema de examen en 2008, y constatado desde que dispusimos del instrumento *PC-aciertos*, de que muchos alumnos con buen nivel de conocimientos manifestaban aversión al riesgo y escogían niveles de confianza por debajo de su dominio de la materia es lo que nos llevó a modificar la tabla original de puntuaciones propuesta por Gardner–Medwin (Tabla 2.17 y 2.18 del capítulo «Metodología»). En la tabla original se penalizaba el error con máximo nivel de seguridad con –4 puntos; nosotros rebajamos esa penalización a –3 con la esperanza de inducir a los alumnos con buen nivel de conocimientos a asumir más riesgos. Tras esta modificación conseguimos un pequeño aumento en la confianza de los estudiantes, pues el IC medio pasó del 1.01 obtenido en

los grupos de alumnos anteriores a 2012, en que aplicábamos la penalización de -4, al 1.08 de los grupos posteriores a 2012, en los que se aplicaba la nueva penalización de -3; es decir, los alumnos se arriesgan ahora un poco más.

Cuando el estudiante tiene un conocimiento realista de su dominio de la materia, es decir, cuando su confianza se ajusta a su conocimiento, sus resultados con nuestro sistema de examen mejoran. Es también algo establecido que las expectativas realistas del alumno se asocian a mejores resultados académicos que las demasiado optimistas o pesimistas (Nicholson, Putwain, Connors y Horny-Atkinson, 2013); pero nosotros no hemos podido constatar ese hecho, al menos en la peculiar forma de valorar la *autoconfianza* que supone el TBC. Como se puso de manifiesto por la ausencia de correlación significativa entre el índice $|PC\text{-aciertos}|$ en valor absoluto y la puntuación por aciertos (r de 0,086, $p > 0.05$), en nuestro estudio no se puede constatar que los alumnos con más conocimientos tengan expectativas de resultados más realistas que los alumnos con menos aciertos en el test. Nuestros estudiantes utilizan con similar habilidad el TBC con independencia del número de aciertos que tengan. Hay que destacar que los resultados señalados en el trabajo de Nicholson, se obtuvieron mediante la comparación de las respuestas de los estudiantes en un test de confianza (el ABC test) y los resultados académicos obtenidos un semestre después. Nuestro sistema de evaluación de la confianza es muy distinto, es coetáneo a la evaluación -teóricamente es, incluso, retrospectivo- y, debido a que condiciona la puntuación del alumno, es pluridimensional, no solo evalúa la confianza sino también las estrategias del alumno a la hora de escoger su opción de seguridad, mediadas por sus objetivos. De acuerdo con nuestros datos, podemos afirmar que, con nuestro sistema de examen, los estudiantes, saben utilizar en igual medida el TBC con independencia de sus niveles, altos o bajos, de conocimiento.

Como hemos visto, con el índice *PC-aciertos*, obtenido en el propio TBC sin que su cálculo suponga una sobrecarga de trabajo para el profesor, tenemos un instrumento potente para conocer cuál ha sido la actitud del estudiante frente al examen; con él sabemos si ha sido excesivamente arriesgado, muy prudente o ha tenido el nivel de confianza acorde con sus conocimientos. Es un instrumento muy fácil de obtener, que puede ser valioso para la tutoría académica. Con una sola cifra, que se calcula simultáneamente al cálculo de la nota del examen, conocemos un rasgo importante de la actitud del alumno frente a las evaluaciones, actitud que, de otra manera, obligaría a realizar un encuesta para conocerla. En cierta manera, podríamos decir que, con la PC, en cada examen realizamos una encuesta a los alumnos sobre su autoconfianza.

Este instrumento para medir la confianza nos ha permitido comprobar un hecho que se daba por supuesto en el TBC, pero que hasta ahora no se había demostrado matemáticamente. Es el hecho de que el TBC premia la sinceridad del alumno al expresar su nivel de confianza, mejorando su puntuación, y castiga, igualmente, su falta de sinceridad (Gardner-Medwin, 2006). Aunque todas las tablas de puntuación de los diversos sistemas de TBC están pensadas para otorgar una penalización en la nota al exceso o defecto de confianza, nosotros hemos podido demostrarlo en el estudio de correlaciones utilizando la PC y el índice TBC-aciertos. La correlación entre el índice *PC-aciertos* en valor absoluto –sin signo negativo- y el *TBC-aciertos*, que expresa la mejora o empeoramiento de la puntuación con el TBC frente a la nota *por aciertos*, es de -0.40, alta y estadísticamente significativa ($p < 0.00001$). Esa correlación negativa significa que cuanto más se aleja –por exceso o por defecto- la PC de la nota del examen tanta menor puntuación recibe el alumno. Y la igualdad entre la puntuación en confianza y la nota en el examen indica que el nivel de confianza del estudiante es acorde al nivel de conocimientos. El alumno que posee una conciencia exacta de cuál es

su nivel de conocimientos, y es sincero al expresarlo, resulta beneficiado en el test basado en la confianza con un aumento de su nota. El insincero o carente de metacognición es penalizado. El TBC tiende a premiar el conocimiento de más calidad, conocimiento que incluye un correcto *conocimiento del propio conocimiento* – metacognición-, que reconoce sus límites, sus fortalezas y debilidades, que evita los errores de la excesiva autoconfianza y la parálisis de la inseguridad; un *conocimiento que se acompaña de una adecuada confianza, ni excesiva ni demasiado escasa*.

Eso hace que este sistema de examen, una vez que el estudiante se familiariza con él, estimule un aprendizaje en el que los meros conocimientos han de acompañarse de seguridad, y esta solo se obtiene de una comprensión profunda de la materia aprendida. Creemos que por este motivo los alumnos encuestados opinaron que el TBC era el sistema de examen que más les forzaba a intentar comprender bien todo lo que aprendían. También creemos que esta puede ser la razón por la que los alumnos con un enfoque de aprendizaje profundo, centrado más en la comprensión que en la mera reproducción de los contenidos, son los que salen más beneficiados en el test basado en la confianza, como vimos que ocurría.

6.12. CONCLUSIÓN FINAL

El TBC fuerza al alumno a la reflexión sobre sus conocimientos, fuerza su metacognición y su autorreflexión; así mismo, estimula el esfuerzo de estudio y comprensión, fomenta los enfoques profundos de aprendizaje, las estrategias de autorregulación y premia, y por lo tanto motiva, la sinceridad del estudiante. Hace al profesor y al alumno conocer mejor la progresión del aprendizaje. Mejora la validez y

eficacia del sistema tradicional de test, basado solo en el número de aciertos y fallos. Además de lo anterior, el TBC proporciona datos sobre la autoconfianza del estudiante - la tendencia a la prudencia o al riesgo en el examen- lo que lo convierte en un instrumento útil para la tutoría académica del alumno y puede permitir posteriores investigaciones en este campo. Cumple, de esta manera, todas las condiciones que se le pueden exigir a un buen sistema de evaluación. Creemos que supone una mejora en todos los aspectos sobre el sistema de exámenes de tipo test convencionales, donde la nota surge solo del número de aciertos y errores.

Pero el TBC no solo es un sistema de evaluación, es un sistema pedagógico amplio que puede actuar en todas las etapas del aprendizaje. El TBC trabaja con los conocimientos y con la confianza. Su uso continuado a lo largo del curso, no solo en el examen final, sino en pequeñas y repetidas evaluaciones así como en autoevaluaciones tutorizadas, tiende a crear en el alumno un hábito de automonitorización continua del aprendizaje (Gardner-Medwin y Curtin, 2007; Nix, 2011); crea un hábito de reflexión, tanto sobre las bases científicas de los conceptos aprehendidos como sobre su propio proceso de aprendizaje; reflexión sobre su propia seguridad e inseguridad, sobre las razones profundas en las que se asientan su confianza y sus dudas. De todo esto se puede seguir un mayor ajuste entre conocimientos y confianza, evitando excesos temerarios de seguridad y faltas de confianza paralizantes; también un mayor conocimiento de sí mismo, de sus potencialidades y sus limitaciones. Al trabajar sobre aprendizaje y autoconfianza, no solo puede contribuir a que el estudiante adquiera el dominio de una materia concreta, también puede contribuir a formarlo mejor como futuro profesional.

Por todo ello, podemos afirmar que el test con puntuación basada en el grado de confianza es un instrumento excelente para el aprendizaje y la evaluación del alumno, mejor que los sistemas más habituales de test, y propugnamos que su utilización en nuestro país se haga más frecuente.

7. BIBLIOGRAFÍA

- Adams, T., Ewen, G. (2009). The importance of confidence in improving educational outcomes, *Proceedings of the 25th Annual Conference of Distance Teaching and Learning*. Board of Regents of the Wisconsin University.
- Ahlgren, A. (1969). Reliability, predictive validity, and personality bias of confidence-weighted scores. *Paper read at the American Educational Research Association Convention, Los Angeles, California, feb5–8, 1969. U. S. Department of Health, Education & Welfare, Office of Education*, document resume 33384, CG 004008.
- Ahlgren, A. (1970). A hand-scoring system for Confidence-Weighted scores. *U. S. Department of Health, Education & Welfare, Office of Education*, document resume 48327, TM 000329, Sep 1970.
- Álvarez, I.M. (2009). Evaluar para contribuir a la autorregulación del aprendizaje. *Electronico Journal of Research in Educational Psychology*, 7(3), 1007-1030.
- Anderson, R. (1982). Computer-based Confidence Testing: Alternatives to Conventional Computer-Based Multiple-Choice Testing. *Journal of Computer-Based Instruction*, 9 (1), 1–9.
- Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman. B
- Balch, W.R. (2007). Effects of Test Expectation on Multiple-Choice Performance and Subjective Ratings. *Teaching of Psychology*, Vol. 34 Issue 4, p 219–225.
- Bandura, A. (1977). Self-efficacy; Toward a unifying theory of behavioural change. *Psychological Reviews*, 84, 191-215.
- Bandura, A. (1986). *Social foundation of thought and action*. London: Englewood Cliffs.

- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.
- Barr, D.A., y Burke, J.R. (2013). Using confidence-based marking in a laboratory setting: A tool for students self-assessment and learning. *The Journal of Chiropractic Education*, 27(1), 21.
- Ben-Shakhar, G., y Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies, *Journal of Educational Measurement*, 12, 428-454.
- Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology*, 48, 266-279.
- Biggs, J.B. (1979). Individual differences in study processes and the quality of learning outcomes. *Higher Education*, 8, 381-394.
- Biggs, J. B., (1982): «Student motivation and study strategies i university and collage of advanced education populations», *Higher Educational Research and Development*, 1, 33-55.
- Biggs, J. B., (1987): *The Study Process Questionnaire (SPQ): Manual* Hawthorn (VIC): Australian Council of Educational Research
- Biggs, J. B, Kember, D., Leung, D. (2001): The revised two-factor Study Process Questionnaire: R-SPQ-2F, *British Journal of Educational Psychology*, 71 (1), 133-149.
- Biggs, J. (1993). What do inventories of students' learning process really measure? A theoretical review and clarification, *British Journal of Educational Psychology*, 83, 3-19.
- Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: the classification of educational goal S: handbook I, cognitive domain* (London, Longman Group).
- Boekaerts, M. y Corno, L. (2005). Self-regulation in the classroom: A perspective on assesment and intervention. *Applied Psychology*, 54(2), 199-231.

- Bong, M. y Skaavik, E.M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1-40.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in Instructional Psychology*, Vol. 1 (pp. 77–165). Hillsdale: Erlbaum.
- Brown, A. (1987). Metacognition, executive control, self–regulation and other Mysterious mechanisms. En F. E. Weinert y R. H. Kluwe (Eds.), *Metacognition, motivation and understanding*, (65–116). Hillsdale, NJ: Erlbaum. B.
- Budescu, D., and Bar-Hillel, M., (1993). To Guess or not to Guess: A Decision–Theoretic View of Formula Scoring, *Journal of Educational Measurement*, 30, (4), 277–291.
- Burton, R. F. (2005). Multiple–choice and true/false test: myths and misapprehensions, *Assessment & Evaluation in Higher Education*, 30(1), pp. 65–72.
- Bush, M. (2001). A multiple choice test that rewards partial knowledge, *Journal of Further and Higher Education*, 25(2), pp. 157–163.
- Chang, S.H., Lin, P.C. y Lin, Z.C. (2007). Measures of Partial Knowledge and Unexpected Responses in Multiple-Choice Tests, *Educational Technology & Society*, 10(4), 95–109.
- Chang, C.Y., Cheng, W.Y. (2008). Science Achievement and Students' Self-Confidence and Interest in Science: A Taiwanese Representative Sample Study, *International Journal of Science Education*, 30 (9), 1183-1200
- Chen, P.P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seven-grade mathematics students. *Learning and Individual Differences*, 14, 79-92.
- Cisar, S.M., Cisar, P., y Pinter, R. (2009). True/false questions analysis using computerized certainty-based marking tests. In *Intelligent Systems and Informatics, 2009. SISY'09 7th International Symposium on* (pp. 171-174). IEEE.
- Collet, L. S. (1971). Elimination scoring: An empirical evaluation, *Journal of Educational Measurement*, 8, 209–214.

- Connelly, L. B. (2004). Assertion–reason assessment in formative and summative tests: Results from two graduate case studies. In *Educational Innovation in Economics and Business*, (pp. 359–378). Springer Netherlands.
- Coombs, C.H. (1953): “On the Use of the Objective Exa”, *Educational and Psychological Measurement*, 13, 308–310.
- Coombs, C. H., Miholland, J. E., y Womer, F. B. (1956). The assessment of partial knowledge. *Educational and Psychological Measurement*, 16, 13–37.
- Cronbach, L.J. (1970). Review of the book “On the theory of achievement test items”. *Psychometrika*, 35, 509-511
- De La Fuente, J., Pichardo, M., Justicia, F. y Berbén, A. (2008). Enfoques de aprendizaje, autorregulación y rendimiento en tres universidades europeas. *Psicothema*, 20(4), 705–711.
- Downing, S.M. (1992). True-False, Alternate-Choice and Multiple Choice items. *Educational Measurement: Issues and Practice*, 11(3), 27-30.
- Dressell, P.L. y Schmid, J (1953).Some Modifications in the Multiple Choice Item, *Educational and Psychological Measurement*, 13, 574–595.
- Ebel, R.L. (1968): Valid confidence testing: Demonstration kit, *Journal of Educational Measurement*, 5 (4), 353–354.
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46, 6-25.
- Ekins, J. M. (1992). *The Development of Study Processes in Distance Learning Students*, Paper presented at a meeting of the Asian Association of Open Universities (Korea, Sept. 1992) ERIC NUMBER, ED355402
- Entwistle, N.; Hanley, M., y Ratcliffe, G. (1979) “Approaches to learning and levels of understanding”, *British Journal of Educational Research*, 5, 99–114.
- Entwistle, N. J. y Ramsden, P., (1983). *Understanding student learning*. London: Croom H. M., (c. 8).

- Feld, L.S. (1980). A test of the hypothesis that Cronbach's alpha coefficient is the same for two tests administered to the same sample. *Psychometrika*, 49, 99-105.
- Fenning, B.E. y May, L.N. (2013). Where there is a will, there is an A: Examining the roles of self-efficacy and self-concept in college students' current educational attainment and career planning. *Social Psychology of Education*, 16, 635-650.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring, *American Psychologist*, 34, 906-911.
- Fong, C.J., y Krause, J.M. (2014). Lost confidence and potential: a mixed methods study of underachieving college students' sources of self-efficacy. *Social Psychology of Education*, 17, 249-268.
- Fransson, A. (1977). On Qualitative Differences in Learning. IV – Effectives of motivational test anxiety on process and outcome. *British Journal of Educational Psychology*, 47, 244-257.
- Frary, R. B. (1982), A Simulation Study of Reliability and Validity of Multiple-Choice Test Scores Under Six Response-Scoring Modes. *Journal of Educational Statistics*, 7, 333-351.
- Gardner W.C. (1969). The use of Confidence Testing in the Academic Instructor Course. *Paper presented at the Annual Conference of the Military Testing Association* (11th, Governors Island, New York, September 15-19)
- Gardner-Medwin, A.R. (1995). Confidence assessment in the teaching of basic science, *Association for Learning Technology Journal* 3:80-85.
- Gardner-Medwin, A. R. (1998). Updating with Confidence: Do your students know what they don't know? *Health Informatics*, 4: 45-46. Disponible en <http://www.ucl.ac.uk/~ucgbarg/tea/ctilap2.htm>.
- Gardner-Medwin, A. R. (2005). Implementing Confidence-Based Marking with your own technology. *Paper presented at the 12th International Conference of the Association for Learning Technology which was held at the University of Manchester, England.*
http://www.Alt.Ac.uk/altc2005/timetable/abstract.Php?abstract_id=593.

- Gardner–Medwin, A.R. (2006a). Confidence-Based Marking towards deeper learning and better exams. En Bryan C. and Clegg K. (Eds.). *Innovative Assessment in Higher Education*. London: Francis and Taylor. Disponible en <http://www.ucl.ac.uk/lapt/innovass6.doc>
- Gardner–Medwin, A.R. (2006 b). Analysis of exams using certainty-based marking. Physiological Society Main Meeting, UCL, *Proceedings of the Physiological Society* series 3, PC64
- Gardner–Medwin, A.R (2007) *Certainty-Based Marking (CBM) for Reflective Learning and Proper Knowledge Assessment*. <http://www.ucl.ac.uk/lapt/REAPcbm.Pdf>
- Gardner–Medwin, A.R. (2011). Reasonable Doubt: Uncertainty in Education, Science and Law. *Proceedings of the British Academy*, 171, 465-483.
- Gardner-Medwin, A.R., y Curtin, N. (2007). Certainty-Based Marking (CBM) for Reflective learning and proper Knowledge Assesment. *Assesment Design for Learner responsibility, 29th-31st May, 2007. Available in <http://ewds.strath.ac.uk/REAP07>*
- Gardner-Medwin, A.R. y Gahan M. (2003). *Formative and Summative Confidence-Based Assessment. Proceedings 7th International Computer–Aided Assessment, Conference, Loughborough, UK, July 2003, 147–155*. Disponible en <http://www.ucl.ac.uk/~ucgbarg/tea/caa03.doc>
- Gargallo López, B., Garfella Esteban, P. R., y Cruz Pérez, P.C. (2006). Enfoques de aprendizaje y rendimiento académico en estudiantes universitarios. *Revista de Orientación Pedagógica*, 58 (3), 327-343.
- Gibbs, G. (1999) Using assessment strategically to change the way students learn, in: S. Brown & A. Glasner (Eds) *Assessment Matters in Higher Education. Choosing and Using Diverse Approaches*, pp. 41–53.
- González Flores, H.A., García Campuzano, D.G. y Vázquez García, F. (2014). Investigación educativa: el aprendizaje autorregulado. Experiencias educativas desde la metacognición y motivación en la formación docente inicial. *Revista iberoamericana de producción Académica y Gestión Educativa*, 1-18.

- González-Pienda, J. A., y otros (coord.), (2008). *Manual de Psicología de la Educación*, Madrid, Ediciones Pirámide.
- Greene, E. B., (1929). Achievement and Confidence on True–False tests of college students. *The Journal of Abnormal and Social Psychology*, Vol 23(4), 467–478.
- Haladyna, T. M. (2004). *Developing and Validating Multiple–Choice Test Items*, (Mahwah, NJ: Lawrence Erlbaum). p. 217.
- Haladyna, T. M. y Downing, S.M. (1993). How many options is enough for a multiple-choice test item?. *Educational and Psychological Measurement*. 53(4),999-1010.
- Haladyna, T. M., Downing, S. M., y Rodríguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15 (3), 309-333.
- Hambelton, R.K., Roberts D. M. & Traub R.R., (1970).A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple choice test, *Journal of Educational Measurement*, 7, 75–90.
- Hassmen, P. y Hunt, D. P., (1994). Human Self–Assessment in Multiple–Choice Testing, *Journal of Educational Measurement*, 31, 149–160.
- Hench, T.L, (2014). Using confidence as feedback in multi-sized learning, *Turkish Online Journal of Educational Technology, TOJET*, 13 (4), 148-156.
- Hevner, Kate (1932). A Method of Correcting for Guessing in True–False Tests and Empirical Evidence in Support of IT', *the Journal of Social Psychology*, 3: 3, 359 — 362.
- Hernández Pina, F. (1993). Concepciones en el estudio de los estudiantes Universitarios, *Revista de Investigación Educativa*, 22, 2º semestre, 117–150.
- Hernández Pina, F., García, M. P., Maquilón, J. J. (2001). «Estudio empírico de los enfoques de aprendizaje de los estudiantes universitarios en función del perfil de su titulación (profundo vs superficial)», *Revista Española de Orientación y Psicopedagogía, REOP*, 12(22), 303–318.

- Hunt, D. P. (2003). "The concept of Knowledge and how to measure it", *Journal Of Intellectual Capital*, 4 (1), 100–113.
- Hutchinson, T. P., (1982). Some theories of performance in multiple choice tests, and their implications for variants of the task. *British Journal of Mathematical and Statistical Psychology*, 35, 71–89.
- Isaacson, R.M., y Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning, *Journal of Scholarship of Teaching and Learning*, 6 (1), 39-55.
- Kansup, W., Hakstian A.R., (1975). A Comparison of Several Methods of Assessing Partial Knowledge in Multiple Choice Tests: I. Scoring Procedures, *Journal of Educational Measurement*, 12, 4, 219–30.
- Karabenick, S. A. (1996). Social influences on metacognition: effects of co-learner questioning on comprehension monitoring. *Journal of Educational Psychology*, 88, 689.
- Kasperski, R. y Katziu, T. (2013). Are Confidence Ratings Test– or Trait–Driven?: individual differences among high, average, and low comprehenders in fourth grade. *Reading Psychology*, 34 (1), 59–84.
- Kelemen, W.L. (2000). Metamemory cues and monitoring accuracy: a cue-utilization approach to judgement of learning, *Journal of Educational Psychology*, 92 (4), 800-810
- Kember, D. y Harper, G. (1987). "Approaches to studying research and its implications for the quality of learning from distance education", *Journal of Distance Education / Revue de l'enseignement a distance*, 2, 2.
- Kember, D., Charlesworth, M., Davies, H., McKay, J., y Sivan, A. (1997). Evaluating the effectiveness of educational innovations, using the Study-Process Questionnaire to show that meaningful learning occurs. *Studies in Educational Evaluation*, 23(2), 141–157.

- Khan K. S., Davies D. A. y Gupta J.K. (2001) Formative self–assessment using multiple True–False questions on the Internet: feedback according to confidence about correct Knowledge. *Medical Teacher*, 23, 158-163
- King P. y Witt P. (2009). Teacher immediacy, Confidence Testing, and the measurement of cognitive learning. *Communication Education*, 58 (1), 110-123.
- Kleitman, S. y Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Journal of Applied Cognitive Psychology*, 15, 321.
- Krathwohl, D. (2002). A revision of Bloom’s Taxonomy: an Overview. *Theory into Practice*, 41 (4), 212-218.
- Kurz, T. B. (1999). A review of scoring algorithms for multiple-choice tests. *Paper presented at Annual Meeting of the Southwest Educational Research Association*, January 21–23, San Antonio, TX.
- Lau. P. N. K., Lau, S. H., Hong, K. S., y Usop, H. (2011). Guessing, Partial Knowledge, and Misconceptions in Multiple–Choice Tests. *Educational Technology & Society*, 14 (4), 99–110.
- Lau, S.H., Hong, K. S., Lau, P.N.K. y Usop, H. (2014). Robustness of number right elimination testing(NRET) scoring method for multiple–choice items in computer adaptive assessment system (CAAS), *Research and Practice in Technology Enhanced Learning*, 9, (2) 283–300.
- Leclerc, D. (1982). «Confidence Marking: Its Use in Testing. *Evaluation in Education: An International Review Series*, v6 n2, 163–287.
- Mansell, W. (2008). Boards explore cheaper option of online multiple–choice tests. *Times Educational Supplement* (10–10–2008), Issue 4809, 22–26.
- Martí, E. (1995). Metacognición: entre la fascinación y el desencanto. *Infancia y aprendizaje*, 72, 9-32.
- Marton, F., y Säljo, R. (1976a). Qualitative differences of learning (I): outcome and process, *British Journal of Educational Psychology*.46, 115–117.

- Marton, F., y Säljö, R. (1976b). On qualitative differences in learning (2): outcome as a function of the learner's conception of the task, *British Journal of Educational Psychology*, 46, 115–127.
- Marsh, H.W., y Craven, R.G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1 (2), 133-166.
- Matheus, J. (2006). *Just Whose Idea Was All This Testing?* The Washington Post, <http://www.washingtonpost.com/wp-dyn/content/article/2006/11/13/AR2006111301007> (Consultado el 10 de junio de 2015).
- Mayer, R.E. (2002). Rote versus meaningful learning. *Theory into Practice*, 41 (4), 226-232.
- Moore, D.A. y Healy, P.J. (2008). The trouble with overconfidence. *Psychological Review*, 115 (2), 502-517.
- Morales Vallejo, P. (2006a). Las pruebas objetivas: normas, modalidades y cuestiones discutidas. (Consultado el 12 de febrero de 2015) <http://www.upcomillas.es/personal/peter/otrosdocumentos/PruebasObjetivas.Pdf>.
- Morales Vallejo, P. (2006b). *Medición de actitudes en psicología y educación*, 3ª edición revisada. Madrid: Universidad Pontificia Comillas.
- Morales Vallejo, P. (2008). *Estadística aplicada a las Ciencias Social Es*. Madrid: Biblioteca de Comillas, Ciencias Social Es. Publicaciones de la Universidad Pontificia de Comillas.
- Morgan, P.J., y Cleave-Hogg, D. (2002). Comparison between medical students' experience, confidence and competence. *Medical education*, 36(6), 534-539.
- Morony, S., Kleitman, S., Lee, Y. P., y Stankov, L. (2013). Predicting Achievement: Confidence versus Self-Efficacy, Anxiety, and Self-Concept in Confucian and European Countries. *International Journal of Educational Research*.58, 79-96.

- Narloch, R., Garbin, Calvin P. and Turnage, K.D. (2006). Benefits of Prelecture Quizzes. *Teaching of Psychology*. Vol. 33 n° 2, 109–112.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, 51, 102–116.
- Nicholson, L., Putwain, D., Connors, L., y Horny-Atkinson, P. (2013). The key to successful achievement as an undergraduate student: confidence and realistic expectations? *Studies in Higher Education*, 38 (2), 285-298.
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education*, 31 (1), 53-64.
- Nix, I., y Wyllie, A. (2011). Exploring design features to enhance computer-based assessment: Learners' views on using a confidence indicator tool and computer-based feedback. *British Journal of Educational Technology*, 42(1), 101-112.
- Osterlind, S. J. (1998). *Construction Test Items, Multiple Choice, Constructed-Response, Performance, and Other Formats*. London: Kluwer Academic Publishers.
- Oh, H. J. (2004, April). Reasoning test takers' guessing strategy and their understanding of formula scoring. *Paper presented at annual meeting of the American Educational Research Association (AERA)*, San Diego, CA.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48, 265.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L, y otros (2002). The role of individual differences in the realism of confidence judgements. *The Journal of General Psychology*, 129, 257-300.
- Panadero, E. y Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales de Psicología*, 30(2), 450-462.
- Papinczak, T. (2009) Are deep strategic learners better suited to PBL? A preliminary study. *Advances in Health Science Education*, 14, 337–353.

- Pérez, V. M. V., Valenzuela, C. M. F., Díaz, M. A., González-Pianda, J.A. y Núñez, P. J. C. (2011). Disposición y enfoques de aprendizaje en estudiantes universitarios de primer año. *Universitas Psychologica*, 10 (2), 441-449.
- Pérez de Albéniz, A., Escolano, E., Pascual, M.T., Lucas, B., Sastre, S. (2015). Metacognición en un proceso de aprendizaje autónomo y cooperativo en el aula universitaria. *Contextos Educativos*, 18, 95-108.
- Pintrich, P.R. (1995). "Understanding self-regulating learning". In P.R. Pintrich (Ed.), *Understanding self-regulated learning* (pp 3-12). San Francisco. CA: Jossey-Bass.
- Pintrich, P.R. (2002). The Role of Metacognitive Knowledge in Learning, Teaching and Assessing. *Theory into Practice*, 41 (4), 219-225.
- Pomplun, M. y Omar, M. H. (1997). Multiple-Mark Items: An Alternative Objective Item Format? *Educational and Psychological Measurement*, 57(6):949-962.
- Ramdass, D. y Zimmerman, B.J. (2008). Effects of self-corrections strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, 1, 18-41.
- Ramdass, D., Zimmerman, B.J. (2011). Developing self-regulation skills: The important role of homework. *Journal of Advanced Academics*, 22(2), 194-218.
- Ramsden, P. y Entwistle, N.J. (1981). Effects of academic departments on students' Approaches to studying. *British Journal of Educational Psychology*, 51, 368-383.
- Recio Saucedo M. A., Cabero Almenara, J. (2005). Enfoques de aprendizaje, Rendimiento académico y satisfacción de los alumnos en formación en entornos virtuales. *Revista de Medios y Educación*, 25, 93-115.
- Reid, F. (1977). An Alternative Scoring Formula for Multiple-Choice and True-False Tests. *Journal of Educational Research*, 70, 335-339.
- Reeve, R. A. y Brown, A.L. (1984), Metacognition reconsidered: Implications for intervention research. *University of Illinois at Urbana-Champaign. Technical report n° 328.2 - 31.*

- Reilly, R.R. (1975), Empirical Option Weighting with a Correction for Guessing. *Educational and Psychological Measurement*, 35, 613–619.
- Rippey R. M. (1978), Interactive confidence test scoring and interpretation. *Educational and Psychological Measurement*, 38 (1), 153–157.
- Rippey R.M., Voytovich A. E. (1983). Linking Knowledge, Realism and Diagnostic Reasoning by Computer–Assisted Confidence Testing. *Journal of Computer – Based Instruction*, 9 (3), 88–97.
- Ritter, L. (2000). The Quest for an Effective Assessment: the evolution and evaluation of a controlled assessment procedure. *Assessment and Evaluation in Higher Education*, Vol. 25 Issue 4, p307–320.
- Rød, J.K., Eiksund S. y Fjær O., (2010). Assessment based on exercise work and Multiple-Choice tests. *Journal of Geography in Higher Education*, 34 (1), pg. 141–153.
- Rodríguez, M.C. (2005). Three options are optimal for Multiple-Choice items: A Meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Rosário, P., Núñez, J., González–Pienda, J., Almeida, L., Soares, S. y Rubio, M. (2005). El aprendizaje escolar examinado desde la perspectiva del «Modelo 3P» de J. Biggs. *Psicothema*, 17 (1), 20–30.
- Rosário, P., Mourão, J., Núñez, J., González-Pienda, J., Solano, P. y Valle, A. (2007). Eficacia de un programa instruccional para la mejora de procesos y estrategias de aprendizaje en la enseñanza superior. *Psicothema*, 19 (3), 422–427.
- Sainsbury, E. J., y Walker, R. A. (2008). Assessment as a vehicle for learning: extending collaboration into testing. *Assessment & Evaluation in Higher Education*, Vol. 33 Issue 2, 103–117.
- Salmerón Pérez, H., Gutiérrez Braojos, C., Salmerón Vílchez, P., Rodríguez Fernández, S. (2011). Metas de logro, estrategias de regulación y rendimiento académico en diferentes estudios universitarios. *Revista de Investigación Educativa*, 29(2), 467-486.

- Sander, P. (2009). Current development in measuring academic behavioural confidence. *Psychology Teaching Review*, 15 (1), 32-44.
- Sander, P. y Sanders, L. (2003). Measuring confidence in academic study: A summary report, *Electronic Journal of Research in Educational Psychology and Psychopedagogy* 1, 1.
- Sander, P. y Sanders, L. (2006). Understanding academic confidence, *Psychology Teaching Review*, 12 (1), 29-42.
- Scouller, K. M. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examinations versus assignment essay. *Higher Education*, 35, 453-472.
- Scouller, K. M. (2000). *The influence of assessment on student learning*, Paper presented at the Annual Conference of the Australian Association for Research in Education, Sydney, Australia. <http://www.aare.edu.au/00pap/sco00195.html> (consultado 26, Feb. 2010).
- Sharma, M.D. y Bewes, J. (2011). Self-Monitoring: Confidence, Academic Achievement and Gender Differences in Physics. *Journal of Learning Design*, 3 (4), 1-13.
- Shoendorfer, N., Emmett, D. (2012). Use of certainty-based marking in a second-year medical student cohort: a pilot study, *Advances in Medical Education and Practice*, Dove Medical Press Ltd, 3, 139-143.
- Sitzmann, T., y Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, 137(3), 421-442.
- Stankov, L. (2000). Complexity, metacognition and fluid intelligence. *Intelligence*, 28(2), 121-143.
- Stankov, L., y Crawford, J. (1997). Self-confidence and performance on cognitive tests. *Intelligence*, 25, 93-109.

- Stankov, L., Lee, J., Luo, W., Hogan, D. J. (2012). Confidence: A better Predictor of Academic Achievement than Self-Efficacy, Self-Concept and Anxiety? *Learning and Individual Differences*, 22 (6), 747-758.
- Stankov, L., Lee, J., y Paek, J. (2009). Realism of confidence judgments. *European Journal of Psychological Assessment*, 25, 123-130.
- Svensson, L. (1977). On qualitative differences in learning III – Study skill and learning. *British Journal of Educational Psychology*, 47, 233–243.
- Swanberg, A. B. y Martinsen, O.L. (2010). Personality, approaches to learning and achievement. *Educational Psychology*, 30 (1), 75–88.
- Swartz, S. M., (2006). Acceptance and Accuracy of Multiple-Choice, Confidence–Level, and Essay-Question Formats for Graduate Students. *Journal of Education for Business*, 81 (4), 215–220.
- Telbys, M. (2010). *Confidence and academic success in higher education*, Online Submission, pp 23.
- Thomson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837-847.
- Tian, X. (2007). Do assessment methods matter? A sensitivity test. *Assessment & Evaluation in High Education*, 32 (4), 387–401.
- Torre Puente, J.C. (2005). *La autoeficacia, la autorregulación y los enfoques de aprendizaje en estudiantes universitarios*. Madrid: Universidad Pontificia Comillas (tesis doctoral)
- Torre Puente, J.C. (2007). *Una triple alianza para un aprendizaje universitario de calidad*. Madrid, Biblioteca Comillas, Educación.
- Torre Puente, J.C. (2012). *Educación y nuevas sociedades. La formación inicial del profesorado de Infantil y Primaria*. Madrid, Publicaciones de la Universidad Pontificia Comillas, Educación, ACISE–FIUC. págs. 245–246.

- Urosa, B. (1995). La adivinación en las pruebas objetivas: Alternativas a la fórmula clásica de corrección. Tesis Doctoral, Universidad Pontificia de Comillas, Madrid, pp 92–128.
- Usher, E.L., y Pajares, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology*, 31, 124-141.
- Usher, E.L., y Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78, 751-796.
- Van Rossum, E. J. y Schenk, S. M. (1984). «The relationship between learning conception, study strategy and learning outcome. *British Journal of Educational Psychology*; 54, 73–83.
- Veenman, V. J., y Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, 15, 159–176.
- Veenman, V. J. y Bernadette, H. A. M. (2006). Van Hout-Wolters. Metacognition and learning: conceptual and methodological considerations *Metacognition Learning (2006) 1: 3–14*
- Vrugt, A. y Oort, F. J. (2008). Metacognition, achievement goal S, study strategies and academic achievement: pathways to achievement. *Metacognition Learning 30: 123–146*
- Wang, M. C., Haertel, G. D., y Walberg, H. J. (1990). What influences learning? A content analysis of review literature, *Journal of Educational Research*, 84 (1), 30–43.
- Wiggins, G. (1990). The case for authentic assessment, *Practical Assessment, Research & Evaluation*, 2 (2). Disponible online en: <http://pareonline.net/getvn.asp?v=2&n=2> (consultado el 5 de marzo de 2015).
- Wilcox, R. (1982). Some New Results on an Answer–Until Correct Scoring Procedure. *Journal of Educational Measurement*, 19 (1), 67–74.

- Williams, J. B. (2006). Assertion–reason Multiple–choice testing as a tool for deep learning: a qualitative analysis. *Assessment and evaluation in higher education*, 31(3), pp. 287–301.
- Winne, P. H. (1996). A metacognitive view of individual differences in self–regulated learning. *Learning and individual differences*, 8(4), 327–353.
- Wolf, L.F., et al., (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education*, 8 (4), 341–351.
- Zimmerman, B. J., y Martinez-Pons, M. (1990). Student differences in self-regulated learning: relating grade, sex, and giftedness to self–efficacy and strategy use. *Journal of Educational Psychology*, 82, 51–59.
- Zimmerman, B. J. (1995). Self–regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist*, 30, 217–221.
- Zimmerman, B. J. (1998). “Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models.” En B. J. Zimmerman y D. H. Schunk (Eds.), *Self-regulated learning: From teaching to self-reflective practice*, (pp. 1-19) New York, NY: Guilford.
- Zimmerman, B.J. (2002). Achieving self-regulation: The trial and triumph of adolescence. En F. Pajares y T. Urban (Eds.), *Academic motivation of adolescents* (vol. 2, pg. 1-27), Greenwich, CT: Information age.
- Zimmerman, B.J. (2013). From cognitive modeling to self-regulation: a social cognitive career path. *Educational Psychologist*, 48(3), 135-147.
- Zimmerman, B.J. y Tsikalas, K. E. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning?, *Educational Psychologist*, 40(4), 267-271.

8. ANEXOS

ANEXO A

FÓRMULA PARA CORREGIR EL ACIERTO POR AZAR EN LOS TEST DE RESPUESTAS MÚLTIPLES.

1.- Fórmula «clásica» de corrección de la adivinación:

- PUNTUACIÓN OBTENIDA = Aciertos – fallos/(N-1), siendo «N» el número de respuestas por cada pregunta.
- NOTA sobre 10 = $10 \times [\text{aciertos} - \text{fallos}/(N-1)] / (\text{número total de preguntas})$

Vamos a demostrar esta fórmula con un desarrollo matemático:

TERMINOLOGÍA:

- Preguntas acertadas en total: **Na**
- Preguntas falladas en total: **Nf**
- Preguntas totales: **N**
- N° de respuestas por pregunta: **n**

Estos son los únicos datos que tenemos del alumno: **Na, Nf, N y n**

El alumno puede acertar preguntas porque se las sepa o por azar, si las ignora. De manera que, el total de preguntas acertadas (**Na**) se puede dividir en dos grupos: **acertadas con conocimiento (Nac)** y **acertadas con ignorancia (Nai)**, o *por azar*.

- Acertadas con conocimiento: **Nac**
- Acertadas con ignorancia (por azar): **Nai**
- **Na = Nac + Nai**

Tabla de símbolos

DEFINICIÓN	Símbolo
Preguntas totales	N
Preguntas acertadas en total	Na
Preguntas falladas	Nf
Número de respuestas por cada pregunta	n
Preguntas acertadas con conocimiento	Nac
Preguntas acertadas por azar	Nai

Suponemos que todas las preguntas falladas por el alumno lo son por ignorancia (es incorrecto, pero el sistema cuenta con eso; por otro lado, es imposible distinguir qué

preguntas son falladas por ignorancia y cuáles por un descuido), luego **Nf** = *preguntas falladas siempre por ignorancia*.

Así pues, hay tres posibles situaciones entre las preguntas respondidas: **Nac**, **Nai** y **Nf**. Conocemos los valores de **Nf** y de la suma de **Nac** y **Nai** (que es el número de acertadas totales, o **Na**).

La puntuación justa para el alumno sería:

$\mathbf{Nac/N}$ (<i>x10, si la damos en los valores habituales de 0 a 10</i>)

Es decir, *el número de respuestas que el estudiante acierta y sabe* –se supone que no falla ninguna sabiéndosela- *partido por el número total de preguntas*. Pero si dividimos el número de respuestas acertadas entre el de respuestas totales no obtenemos esa nota, sino una nota superior, que sería: $\mathbf{Na/N} = (\mathbf{Nac} + \mathbf{Nai})/\mathbf{N}$

Por lo tanto, el estudiante acierta más preguntas de las que se sabe porque acierta algunas por adivinación o azar.

$(\mathbf{Nac} + \mathbf{Nai})/\mathbf{N}$ será siempre superior a $\mathbf{Nac/N}$, salvo que el alumno se sepa todas las preguntas o no acierte ninguna por azar ($\mathbf{Nai} = 0$).

¿Cómo podemos saber cuáles son las preguntas acertadas con conocimiento, **Nac**? Serán iguales a $(\mathbf{Na} - \mathbf{Nai})$. Como conocemos el número de aciertos (**Na**), solo necesitamos conocer el número de aciertos por azar (**Nai**) para conocer los aciertos por conocimiento (**Nac**).

¿Cómo podemos saber cuál es el número de preguntas acertadas por azar (Nai**)?**

Hay que partir del número de preguntas falladas (**Nf**).

El alumno falla todas las preguntas porque las ignora (es una suposición previa del sistema, aunque sepamos que no siempre es así) pero hay otras que ignora y acierta.

Luego podemos llamar «**I**» a las preguntas que el alumno ignora. Así:

$\mathbf{I} = \mathbf{Nai} + \mathbf{Nf}$ (o sea, que el número de preguntas que el alumno ignora es igual a las preguntas que ha acertado por puro azar –**Nai**– más las que ha fallado –**Nf**–).

Tabla de símbolos completa

DEFINICIÓN	Símbolo
Preguntas totales	N
Preguntas acertadas	Na
Preguntas falladas	Nf
Número de respuestas por cada pregunta	n
Preguntas acertadas con conocimiento	Nac
Preguntas acertadas por azar	Nai
Preguntas que el alumno ignora, pero responde	I

Por azar la posibilidad de acertar una pregunta que se ignora es $1/n$, donde **n** es el número de respuestas. Así, por ejemplo, si hay 2 respuestas, la probabilidad de acertar por azar es de $1/2$ (o un 50%, que es otra manera de expresarlo), si hay tres respuestas la probabilidad de acertar por azar es de $1/3$ (33,33%), si hay 4 respuestas la probabilidad es de $1/4$ (25%), si hay 5, $1/5$ (20%), etc.

Por lo tanto, el número de respuestas acertadas por azar tendería a ser: **I / n**

Es decir, el número de preguntas respondidas al azar partido por el número de respuestas (este es el *valor más probable*).

Por lo tanto: **Nai = I / n**

E igualmente, el número de preguntas falladas será:

$$Nf = I - Nai \rightarrow Nf = I - I/n \rightarrow Nf = I \times n/n - I/n \rightarrow Nf = (I \times n - I)/n \rightarrow Nf = I \times (n-1)/n$$

Luego : **Nf = I · (n-1) / n**

Tenemos, por lo tanto:

$$Nai = I/n$$

$$Nf = I \cdot (n-1)/n. \text{ Destaco esta última fórmula porque todo va a}$$

$$\text{girar en torno a ella: } \underline{Nf = I \cdot (N-1)/N}$$

Tenemos pues: Nai = I/n

Si multiplicamos y dividimos por la misma cantidad el total no se altera,

$$\text{así que: } \underline{Nai = I/n \cdot (n-1)/(n-1)}$$

o, lo que es lo mismo: $N_{ai} = I \cdot (n-1) / n(n-1) \rightarrow N_{ai} = [I(n-1)/n] \cdot 1/(n-1)$

$$\underline{N_{ai} = [I \cdot (n-1)/n] \cdot 1/(n-1)}$$

Como $N_f = I \cdot (n-1)/n$

Por lo tanto, $N_{ai} = (N_f) \cdot 1/(n-1) \rightarrow N_{ai} = N_f / (n-1)$

Así pues, ya sabemos cuál es el número de las preguntas acertadas por azar:

$$N_{ai} = N_f/(n-1)$$

Preguntas acertadas por azar = preguntas falladas/(número de respuestas-1)

Luego...

Las preguntas acertadas con conocimiento son: $N_{ac} = N_a - N_{ai}; N_{ai} = N_f/(n-1)$

$$N_{ac} = N_a - N_f/(n-1)$$

Quedamos en que la nota justa era: N_{ac}/N (y $\times 10$, si se puntúa de 0 a 10)

Resultando que:

$$\text{La puntuación justa es } \rightarrow 10 \cdot [N_a - F/(N-1)]/N$$

Que resulta ser la fórmula clásica de corrección de la adivinación

2.- ¿POR QUÉ EL PROCEDIMIENTO CLÁSICO DE CORRECCIÓN DEL AZAR ES INCORRECTO?

El sistema funcionaría bien si el alumno respondiera a todas las preguntas, las supiera o no. O si dejara en blanco aquellas de las que no tiene absolutamente ningún conocimiento.

Pero existe el «**conocimiento parcial**», que es en los exámenes de respuestas múltiples es la situación que se da cuando el sujeto puede rechazar con probabilidad

alguna o algunas de las respuestas. El conocimiento parcial eleva la probabilidad de acertar una pregunta por encima del simple azar.

Podemos definir el «*conocimiento parcial*» como la situación en la que la probabilidad de acertar una pregunta de respuesta múltiple es mayor que respondiendo al azar; igualmente, podemos definir como «ausencia de conocimiento» la situación que se da cuando la probabilidad de acertar una pregunta de respuesta múltiple es exactamente igual a la de responder al azar.

Por lo tanto, en un examen de respuesta múltiple (test), siendo N el número de respuestas por cada pregunta:

- Conocimiento cierto
 - o Probabilidad de acertar = 1
 - o Probabilidad de fallar = 0
- Ausencia de conocimiento
 - o Probabilidad de acertar = $1/n$
 - o Probabilidad de fallar = $1 - 1/n = (n-1)/n$
- Conocimiento parcial
 - o Probabilidad de acertar $> 1/n$
 - o Probabilidad de fallar $< 1 - 1/n$ o $< (n-1)/n$

La fórmula clásica de corrección de la adivinación es:

$$\text{Puntuación} = \text{Aciertos} - \text{fallos}/(n-1)$$

(siendo N el número de respuestas por cada pregunta, como dijimos)

Esa fórmula neutraliza la probabilidad de acertar por adivinación, que era, como hemos señalado más arriba, $1/(n^\circ \text{ de respuestas})$ en cada pregunta respondida al azar.

Si el alumno responde a preguntas con conocimiento parcial, la puntuación obtenida por estas, cuya posibilidad de acierto era $> 1/n$, tenderá a no ser completamente neutralizada por la penalización. Acertará más que $1/n$, fallará menos que $(n-1)/n$, es decir, su número de aciertos será superior, su número de fallos será inferior. Por lo tanto su nota tenderá a ser superior a la del estudiante que responda solo a las preguntas con conocimiento completo.

Así que el consejo que hay que darle a los alumnos que se enfrenten a un examen de tipo test con penalización por respuesta fallada no es el habitual de que «respondan solo cuando estén seguros» sino el de «solamente dejen de contestar aquellas preguntas de las que no sepan absolutamente nada, de manera que tendrían que escoger la respuesta totalmente al azar». Cualquier conocimiento parcial, por pequeño que sea, aumenta la probabilidad de acertar.

ANEXO B

ABC TEST (Sander & Sanders, 2003; 2006)

¿Hasta qué punto confías en que serás capaz de...?:

1. Estudiar efectivamente por ti mismo en tu entorno privado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
2. Producir tu mejor trabajo bajo condiciones de examen.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
3. Responder a preguntas de un profesor ante un tribunal de profesores.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
4. Llevar adelante tu trabajo hasta conseguir tus metas del curso.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
5. Hacer una presentación ante un pequeño grupo de compañeros.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
6. Asistir a la mayoría de las clases.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
7. Conseguir buenos resultados en tu trabajo.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
8. Acoplarte bien a un productivo debate académico con tus pares.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
9. Hacer al profesor preguntas acerca de las materias que está enseñando en una tutoría.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
10. Hacer al profesor preguntas acerca de las materias que está enseñando durante una clase.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
11. Entender la materia impartida y discutida contigo por los profesores.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
12. Seguir los temas y debates en las clases.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
13. Prepararte exhaustivamente para tutoriales.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
14. Leer los materiales de preparación recomendados.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
15. Desarrollar el trabajo del curso cumpliendo los estándares.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
16. Escribir en un estilo académico apropiado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
17. Pedir ayuda si no entiendes algo.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
18. Ir al día en las lecturas.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
19. Aprovechar la mayoría de las oportunidades de estudio en el grado universitario.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
20. Aprobar las evaluaciones a la primera.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
21. Planificar adecuados programas de revisión.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
22. Permanecer siempre suficientemente motivado.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
23. Hacer tu mejor trabajo en los deberes.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado
24. Asistir a las tutorías.	Nada confiado <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Muy confiado

ANEXO C

Otras escalas de autoconfianza

Subescala de confianza de Telbys (Telbys, M. (2010). Confidence and academic success

«I feel confident that I will achieve academic success based on my current situation»	Strongly disengage					Strongly agree
I have a good understanding of science I have previously studied	1	2	3	4	5	6
I have a good grasp of basic ideas in chemistry such as chemical reactions, atoms and molecules	1	2	3	4	5	6
I have a good understanding of physical sciences	1	2	3	4	5	6
I can understand scientific instruments and their relationship to corresponding formulas	1	2	3	4	5	6

ANEXO D

FORMATO del examen de tipo TBC usado por nosotros.

Fisiopatología del paciente crítico y SV (mayo de 2014)

Nombre y apellidos:

El examen consiste en un test de 45 preguntas con tres respuestas cada, sólo una correcta. Tras cada pregunta el alumno debe poner su grado de confianza de haberla acertado: “totalmente seguro”, “bastante seguro” o “no seguro”.

TABLA de PUNTOS	Totalmente seguro	Bastante seguro	No seguro
Correcta	+ 3	+ 2	+ 1
Incorrecta	-3	-1	0

1.- Una de las siguientes manifestaciones de demencia es propia solo de las etapas avanzadas, ¿cuál?

- A.- Memoria para recuerdos recientes.
B.- Reflejo de chupeteo y prensión palmar.
C.- Pérdida de capacidad de cálculo.

RESPUESTA CORRECTA:

Totalmente seguro

Bastante seguro

No estoy seguro

2.- ¿Cuál de las siguientes combinaciones de 3 síntomas del SNC sería con seguridad una demencia, si fueran persistentes en el tiempo?

- A) Afasia, desorientación, alteración de personalidad.
B) Alteración de la personalidad, apraxia motora, ataxia.
C) Asomatognosia, afasia, apraxia motora.

RESPUESTA CORRECTA:

Totalmente seguro

Bastante seguro

No estoy seguro

3.- ¿En qué parte del cerebro se regulan la vigilia y el sueño?

- A) Sustancia reticular del tronco encefálico.
B) Hipotálamo y núcleos de la base.
C) Área límbica de la corteza cerebral.

RESPUESTA CORRECTA:

Totalmente seguro

Bastante seguro

No estoy seguro

4.- Un paciente que tiene un infarto cerebral izquierdo con hemianopsia contralateral, ¿qué tipo de afectación visual tendrá?

- A) Pérdida del campo izquierdo de ambos ojos.
B) Pérdida del campo izquierdo del ojo derecho y del campo derecho del ojo izquierdo.
C) Pérdida del campo derecho de ambos ojos.

RESPUESTA CORRECTA:

Totalmente seguro

Bastante seguro

No estoy seguro

ANEXO E
ENCUESTA sobre sistemas de evaluación (1).

- EDAD: SEXO:
- ¿Trabajas actualmente?:
- Nota obtenida en el examen de diciembre (letra verde):
- Número de aciertos en el examen de diciembre (letra roja):

A continuación, elige una respuesta de 6 a 1 según estés + ó - de acuerdo con el enunciado de la pregunta, según la siguiente escala:

SI, TOTALMENTE DE ACUERDO						NO, TOTALMENTE EN DESCUEDO
6	5	4	3	2	1	

¿En qué medida te gustan los siguientes sistemas de evaluación?:

2) Test convencional, sin valorar el grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

3) Test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

4) Exámenes de tipo desarrollar un tema

6	5	4	3	2	1
<input type="checkbox"/>					

¿En qué medida cada tipo de examen te obliga a dedicar más tiempo al estudio?

5) Test tradicional, sin indicar grado de seguridad.

6	5	4	3	2	1
<input type="checkbox"/>					

6) Test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

7) Examen tipo desarrollar un tema.

6	5	4	3	2	1
<input type="checkbox"/>					

ANEXO E (2)
ENCUESTA sobre sistemas de evaluación.

¿En qué medida cada tipo de examen te obliga a entender bien todo lo que estás estudiando?:

8) Test tradicional, sin indicar grado de seguridad.

6	5	4	3	2	1
<input type="checkbox"/>					

9) Test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

10) Examen tipo desarrollar un tema abierto.

6	5	4	3	2	1
<input type="checkbox"/>					

¿En qué medida cada tipo de examen te obliga a aprender más datos de memoria (memorizar más)?

11) Test tradicional, sin indicar grado de seguridad.

6	5	4	3	2	1
<input type="checkbox"/>					

12) Test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

13) Examen tipo desarrollar un tema abierto.

6	5	4	3	2	1
<input type="checkbox"/>					

¿Qué tipo de examen crees que le permite al profesor conocer mejor lo que el alumno sabe?

15) Test tradicional, sin indicar grado de seguridad.

6	5	4	3	2	1
<input type="checkbox"/>					

16) Test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

17) Examen tipo desarrollar un tema abierto.

6	5	4	3	2	1
<input type="checkbox"/>					

ANEXO E(3)

ENCUESTA sobre sistemas de evaluación

ANEXO E (4)
ENCUESTA sobre sistemas de evaluación

¿Te sientes muy estresado antes de hacer un examen?

26) Si es un test tradicional, sin indicar grado de seguridad.

6	5	4	3	2	1
<input type="checkbox"/>					

27) Si es un test teniendo que indicar mi grado de seguridad

6	5	4	3	2	1
<input type="checkbox"/>					

28) Si es un examen tipo desarrollar un tema abierto.

6	5	4	3	2	1
<input type="checkbox"/>					

Respecto a tu forma de estudiar, ¿responden a tu situación las siguientes frases?:

29) Me doy cuenta de que estudiar me proporciona con frecuencia un sentimiento de profunda satisfacción personal.

6	5	4	3	2	1
<input type="checkbox"/>					

30) Al elaborar o estudiar un tema, no me encuentro satisfecho hasta que me he formado mis propias conclusiones sobre él.

6	5	4	3	2	1
<input type="checkbox"/>					

31) Mi objetivo es aprobar el curso haciendo el mínimo trabajo posible.

6	5	4	3	2	1
<input type="checkbox"/>					

32) Sólo estudio seriamente lo que se da en las clases o lo que está en los programas detallados de las asignaturas.

6	5	4	3	2	1
<input type="checkbox"/>					

33) Me parece que cualquier tema puede llegar a ser muy interesante una vez que te metes a fondo en él.

6	5	4	3	2	1
<input type="checkbox"/>					

ANEXO E (6)
ENCUESTA sobre sistemas de evaluación

42) Empleo bastante de mi tiempo libre en buscar más información sobre temas interesantes que se han discutido en las diferentes clases.

6	5	4	3	2	1
<input type="checkbox"/>					

43) Me parece que no ayuda estudiar los temas en profundidad. Confunde y hace perder el tiempo cuando todo lo que se necesita es un conocimiento por encima de los temas.

6	5	4	3	2	1
<input type="checkbox"/>					

44) Creo que los profesores no deberían esperar que los alumnos dedicaran mucho tiempo a estudiar cosas que no van a caer en el examen.

6	5	4	3	2	1
<input type="checkbox"/>					

45) Voy a la mayoría de las clases con preguntas a las que desearía encontrar respuesta.

6	5	4	3	2	1
<input type="checkbox"/>					

46) Es muy importante para mí echar un vistazo a la mayoría de las lecturas recomendadas que tienen que ver con las clases.

6	5	4	3	2	1
<input type="checkbox"/>					

47) No le encuentro sentido a aprender contenidos que probablemente no caerán en el examen.

6	5	4	3	2	1
<input type="checkbox"/>					

48) Me parece que la mejor manera de aprobar es recordar las respuestas de las preguntas que con más probabilidad pueden caer en el examen.

6	5	4	3	2	1
<input type="checkbox"/>					

49) Mientras estudio, me doy cuenta de si voy cumpliendo o no mis objetivos.

6	5	4	3	2	1
<input type="checkbox"/>					

ANEXO E (7)
ENCUESTA sobre sistemas de evaluación

50) No utilizo siempre los mismos modos para trabajar/aprender los temas, sé cambiar de estrategia.

6	5	4	3	2	1
<input type="checkbox"/>					

51) Cuando estoy estudiando una asignatura o módulo, sé exactamente qué quiero conseguir o aprender.

6	5	4	3	2	1
<input type="checkbox"/>					

52) Cuando estoy estudiando, ante las dificultades, hago un esfuerzo mayor o cambio mi modo de estudiar.

6	5	4	3	2	1
<input type="checkbox"/>					

53) Cuando estoy trabajando una asignatura, intento identificar las ideas y conceptos que no comprendo.

6	5	4	3	2	1
<input type="checkbox"/>					

ANEXO F (1)

Estadísticos descriptivos de los ítems de la encuesta:

Edad, trabajo, género

EDAD	EDAD	Trabajo	Trabajo	sexo: 1 V	sexo: 1 V
Media	21,69	Media	0,304	Media	0,203
DS	3,709	DS	0,4596	DS	0,4029
Error típico	0,23	Error típico	0,03	Error típico	0,02
Mediana	20,00	Mediana	0,00	Mediana	0,00
Varianza	13,75	Varianza	0,21	Varianza	0,16
Curtosis	7,67	Curtosis	-1,27	Curtosis	0,21
Asimetría	2,35	Asimetría	0,86	Asimetría	1,49
Rango	24	Rango	1	Rango	1
Mínimo	19	Mínimo	0	Mínimo	0
Máximo	43	Máximo	1	Máximo	1
Cuenta	261	Cuenta	252	Cuenta	271

Nota en el TBC, nota según número de aciertos, TBC-aciertos

Nota TBC	Nota TBC	Nota-aciertos	Nota-aciertos	TBC-aciertos	TBC-aciertos
Media	6,16	Media	6,22	Media	-0,04
Error típico	0,119	Error típico	0,115	Error típico	0,0359
Mediana	6,305	Mediana	6,441	Mediana	-0,022
DS	1,9540	DS	1,8783	DS	0,5898
Varianza	3,818	Varianza	3,528	Varianza	0,348
Curtosis	-0,119	Curtosis	0,264	Curtosis	-0,458
Asimetría	-0,359	Asimetría	-0,628	Asimetría	-0,151
Rango	10,47	Rango	10	Rango	3,040
Mínimo	0,4	Mínimo	0,0	Mínimo	-1,8
Máximo	10,87	Máximo	10	Máximo	1,3
Cuenta	270	Cuenta	269	Cuenta	270

PUNTUACIÓN EN CONFIANZA, IC (alumnos de 2014)

<i>Confianza</i>	<i>Confianza</i>	IC	IC
Media	6,11	Media	1,08
Error típico	0,140	Error típico	0,0497
Mediana	6,25	Mediana	0,9816
Desviación estándar	1,6519	Desviación estándar	0,5718
Varianza de la muestra	2,729	Varianza de la muestra	0,327
Curtosis	1,190	Curtosis	42,885
Asimetría	-0,570	Asimetría	5,774
Rango	9,40	Rango	5,38
Mínimo	0	Mínimo	0,587
Máximo	9,4	Máximo	5,97
Cuenta	139	Cuenta	139

ANEXO F (2)
Estadísticos descriptivos de los ítems de la encuesta

Ítems 2, 3, 4

	Ítem 2		Ítem 3		Ítem 4
Media	4,83	Media	3,68	Media	2,56
Error típico	0,065	Error típico	0,084	Error típico	0,093
Mediana	5	Mediana	4	Mediana	2
DS	1,0665	DS	1,3916	DS	1,5283
Varianza	1,137	Varianza	1,937	Varianza	2,336
Curtosis	-0,049	Curtosis	-0,733	Curtosis	-0,734
Asimetría	-0,687	Asimetría	-0,248	Asimetría	0,652
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

Ítems 5, 6, 7

	Ítem 5		Ítem 6		Ítem 7
Media	4,28	Media	5,02	Media	5,46
Error típico	0,059	Error típico	0,051	Error típico	0,058
Mediana	4	Mediana	5	Mediana	6
DS	0,9734	DS	0,8414	DS	0,9519
Varianza	0,948	Varianza	0,708	Varianza	0,906
Curtosis	0,747	Curtosis	1,665	Curtosis	5,321
Asimetría	-0,561	Asimetría	-0,903	Asimetría	-2,172
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (3)
Estadísticos descriptivos de los ítems de la encuesta

Ítems 8, 9, 10

	Ítem 8		Ítem 9		Ítem 10
Media	4,69372694	Media	5,36531365	Media	4,58518519
Error típico	0,06504583	Error típico	0,04659133	Error típico	0,08522558
Mediana	5	Mediana	6	Mediana	5
DS	1,07078949	DS	0,76699015	DS	1,40039923
Varianza	1,14659013	Varianza	0,58827388	Varianza	1,961118
Curtosis	0,53250207	Curtosis	1,86516757	Curtosis	-0,09661961
Asimetría	-0,8046456	Asimetría	-1,22625246	Asimetría	-0,85945304
Rango	5	Rango	4	Rango	5
Mínimo	1	Mínimo	2	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	271	Cuenta	271	Cuenta	270

Ítems 11, 12, 13

	Ítem 11		Ítem 12		Ítem 13
Media	4,26568266	Media	4,79335793	Media	5,27675277
Error típico	0,06454928	Error típico	0,06555323	Error típico	0,06276438
Mediana	4	Mediana	5	Mediana	6
DS	1,06261531	DS	1,0791424	DS	1,03323215
Varianza	1,12915129	Varianza	1,16454831	Varianza	1,06756868
Curtosis	0,04612965	Curtosis	0,83067092	Curtosis	1,28888652
Asimetría	-0,36097623	Asimetría	-0,95213328	Asimetría	-1,38710067
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	271	Cuenta	271	Cuenta	271

ANEXO F (4)
Estadísticos descriptivos de los ítems de la encuesta

Ítems 15, 16, 17

	Ítem 15		Ítem 16		Ítem 17
Media	3,96666667	Media	5,04074074	Media	4,68888889
Error típico	0,07047175	Error típico	0,0605805	Error típico	0,08460343
Mediana	4	Mediana	5	Mediana	5
DS	1,15796899	DS	0,99543914	DS	1,39017619
Varianza	1,34089219	Varianza	0,99089908	Varianza	1,93258984
Curtosis	-0,25186752	Curtosis	1,86648894	Curtosis	0,44248522
Asimetría	-0,26754433	Asimetría	-1,24386208	Asimetría	-1,06873585
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	270	Cuenta	270	Cuenta	270

Ítems 18, 19, 20

	Ítem 18		Ítem 19		Ítem 20
Media	4,05514706	Media	4,94852941	Media	4,44117647
Error típico	0,06892426	Error típico	0,06133237	Error típico	0,08359096
Mediana	4	Mediana	5	Mediana	5
DS	1,13672799	DS	1,01151928	DS	1,37861745
Varianza	1,29215053	Varianza	1,02317126	Varianza	1,90058606
Curtosis	-0,23931141	Curtosis	0,86487025	Curtosis	-0,27140498
Asimetría	-0,32136367	Asimetría	-0,97391103	Asimetría	-0,68532778
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (5)

Estadísticos descriptivos de los ítems de la encuesta

Ítems 21, 22, 23

	Ítem 21		Ítem 22		Ítem 23
Media	4,38602941	Media	5,07352941	Media	4,48161765
Error típico	0,06855858	Error típico	0,05318811	Error típico	0,08462722
Mediana	5	Mediana	5	Mediana	5
DS	1,13069699	DS	0,87720079	DS	1,39570782
Varianza	1,27847569	Varianza	0,76948122	Varianza	1,94800033
Curtosis	0,58159326	Curtosis	1,92415114	Curtosis	0,15442903
Asimetría	-0,77117625	Asimetría	-1,10205094	Asimetría	-0,88070346
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

Ítems 26, 27, 28

	Ítem 26		Ítem 27		Ítem 28
Media	4,40073529	Media	4,87132353	Media	5,04411765
Error típico	0,08757457	Error típico	0,07921092	Error típico	0,0784342
Mediana	5	Mediana	5	Mediana	5
DS	1,44431683	DS	1,30637992	DS	1,29356995
Varianza	2,08605112	Varianza	1,7066285	Varianza	1,6733232
Curtosis	-0,2397395	Curtosis	1,47901134	Curtosis	1,92989812
Asimetría	-0,75525857	Asimetría	-1,39037719	Asimetría	-1,58692933
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (6)

Estadísticos descriptivos de los ítems de la encuesta

Ítems 14, 24, 25

	Ítem 14		Ítem 24		Ítem 25
Media	4,59558824	Media	4,11439114	Media	2,61029412
Error típico	0,09328981	Error típico	0,08490352	Error típico	0,08356112
Mediana	5	Mediana	4	Mediana	2
DS	1,53857497	DS	1,3976884	DS	1,37812533
Varianza	2,36721294	Varianza	1,95353287	Varianza	1,89922943
Curtosis	-0,13626689	Curtosis	-0,48916913	Curtosis	-0,47780808
Asimetría	-0,96523679	Asimetría	-0,53351585	Asimetría	0,55650686
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	271	Cuenta	272

Ítems de la escala R-SPQ-2F (enfoques de aprendizaje)

Ítems 29, 30, 31

	Ítem 29		Ítem 30		Ítem 31
Media	4,66	Media	4,61	Media	1,8860
Error típico	0,06919041	Error típico	0,0638	Error típico	0,0635
Mediana	5	Mediana	5	Mediana	2
DS	1,14111739	DS	1,0530	DS	1,04769
Varianza	1,3021489	Varianza	1,1087	Varianza	1,09766
Curtosis	0,35885151	Curtosis	-0,3307	Curtosis	0,44776
Asimetría	-0,80907554	Asimetría	-0,3992	Asimetría	1,06383
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (7)

Estadísticos descriptivos de los ítems de la encuesta

Ítems 32, 33, 34

	Ítem 32		Ítem 33		Ítem 34
Media	3,6801	Media	4,61764706	Media	3,4154
Error típico	0,0806	Error típico	0,07201058	Error típico	0,0697
Mediana	4	Mediana	5	Mediana	3
DS	1,3297	DS	1,18762887	DS	1,1495
Varianza	1,7682	Varianza	1,41046234	Varianza	1,3212
Curtosis	-0,7736	Curtosis	-0,09364618	Curtosis	-0,4907
Asimetría	-0,2882	Asimetría	-0,68872307	Asimetría	0,2674
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

Ítems 35, 36, 37

	Ítem 35		Ítem 36		Ítem 37
Media	3,10294118	Media	2,3309	Media	3,58
Error típico	0,08179112	Error típico	0,0760	Error típico	0,0839
Mediana	3	Mediana	2	Mediana	4
DS	1,34893377	DS	1,2541	DS	1,3831
Varianza	1,81962231	Varianza	1,5728	Varianza	1,9128
Curtosis	-0,84550718	Curtosis	-0,2976	Curtosis	-0,7790
Asimetría	0,07484106	Asimetría	0,7333	Asimetría	0,0998
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (8)

Estadísticos descriptivos de los ítems de la encuesta

Ítems 38, 39, 40

	Ítem 38		Ítem 39		Ítem 40
Media	4,3653	Media	2,3824	Media	2,50
Error típico	0,0727	Error típico	0,0764	Error típico	0,0697
Mediana	5	Mediana	2	Mediana	2
DS	1,1970	DS	1,2600	DS	1,1494
Varianza	1,4327	Varianza	1,5876	Varianza	1,3210
Curtosis	-0,5008	Curtosis	0,3495	Curtosis	-0,0805
Asimetría	-0,4355	Asimetría	0,9316	Asimetría	0,6243
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	271	Cuenta	272	Cuenta	272

Ítems 41, 42, 43

	Ítem 41		Ítem 42		Ítem 43
Media	4,482	Media	2,7684	Media	1,9265
Error típico	0,0563	Error típico	0,0680	Error típico	0,0574
Mediana	4	Mediana	3	Mediana	2
DS	0,9290	DS	1,1208	DS	0,9460
Varianza	0,8631	Varianza	1,2561	Varianza	0,8949
Curtosis	-0,4938	Curtosis	-0,2497	Curtosis	1,4287
Asimetría	-0,1268	Asimetría	0,2612	Asimetría	1,1223
Rango	4	Rango	5	Rango	5
Mínimo	2	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

ANEXO F (9)

Estadísticos descriptivos de los ítems de la encuesta

Ítems 44, 45, 46

	Ítem 44		Ítem 45		Ítem 46
Media	2,9557	Media	3,1765	Media	2,9706
Error típico	0,0873	Error típico	0,0800	Error típico	0,0727
Mediana	3	Mediana	3	Mediana	3
DS	1,4369	DS	1,3191	DS	1,1993
Varianza	2,0647	Varianza	1,7400	Varianza	1,4382
Curtosis	-0,8681	Curtosis	-0,5266	Curtosis	-0,3222
Asimetría	0,3573	Asimetría	0,4679	Asimetría	0,3414
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	271	Cuenta	272	Cuenta	272

Ítems 47, 48

	Ítem 47		Ítem 48
Media	2,2059	Media	2,5478
Error típico	0,0678	Error típico	0,0793
Mediana	2	Mediana	2
DS	1,1176	DS	1,30763
Varianza	1,2490	Varianza	1,70988
Curtosis	0,2895	Curtosis	-0,58416
Asimetría	0,8324	Asimetría	0,54330
Rango	5	Rango	5
Mínimo	1	Mínimo	1
Máximo	6	Máximo	6
Cuenta	272	Cuenta	272

ANEXO F (10)

Estadísticos descriptivos de los ítems de la encuesta

ÍTEMS DE AUTORREGULACIÓN**Ítems 49, 50, 51**

	Ítem 49		Ítem 50		Ítem 51
Media	4,85294118	Media	3,67279412	Media	4,39338235
Error típico	0,06075881	Error típico	0,08900045	Error típico	0,06840417
Mediana	5	Mediana	4	Mediana	5
DS	1,00205996	DS	1,46783305	DS	1,12815051
Varianza	1,00412416	Varianza	2,15453386	Varianza	1,27272357
Curtosis	1,07903726	Curtosis	-0,92119796	Curtosis	0,18916651
Asimetría	-0,8759014	Asimetría	-0,09677015	Asimetría	-0,61774044
Rango	5	Rango	5	Rango	5
Mínimo	1	Mínimo	1	Mínimo	1
Máximo	6	Máximo	6	Máximo	6
Cuenta	272	Cuenta	272	Cuenta	272

Ítems 52, 53

	Ítem 52		Ítem 53
Media	4,80514706	Media	5,09558824
Error típico	0,06214467	Error típico	0,0545711
Mediana	5	Mediana	5
DS	1,02491615	DS	0,90000965
Varianza	1,05045311	Varianza	0,81001736
Curtosis	1,05701481	Curtosis	0,55974104
Asimetría	-0,92735081	Asimetría	-0,89329587
Rango	5	Rango	4
Mínimo	1	Mínimo	2
Máximo	6	Máximo	6
Cuenta	272	Cuenta	272

ANEXO G (1)

Cuadro de correlaciones de los ítems de la encuesta

	EDAD	trabajo	Nota CBT	Aciertos	CBT -aciertos
EDAD	1,0000000				
trabajo	0,3407889	1,0000000			
Nota CBT	0,0379685	-0,1470821	1,0000000		
Aciertos	0,0392793	-0,1130792	0,9521387	1,0000000	
CBT –aciert.	-0,0016638	-0,0675211	0,2203438	-0,0758701	1,0000000
2	0,0057670	-0,0208700	-0,0687409	-0,0451711	-0,0779294
3	0,1611667	0,0997378	0,3489647	0,3053080	0,1665250
4	-0,0074334	0,0598611	-0,0058482	-0,0344334	0,0755156
5	-0,0617843	-0,0744435	0,0619115	0,0507868	0,0325658
6	-0,0611513	-0,0030987	-0,0187283	-0,0457529	0,1286897
7	-0,1352975	-0,0184868	0,0908505	0,0941766	0,0242374
8	0,0300303	-0,0135269	-0,0184254	-0,0050506	-0,0700819
9	0,0821153	-0,0014017	0,0905286	0,0465939	0,0827951
10	-0,0114232	-0,0902253	0,0944034	0,0858543	0,0801223
11	-0,1094046	-0,0007634	-0,0177903	-0,0230988	-0,0323713
12	-0,0986274	0,0531247	-0,0216904	-0,0620203	0,1328137
13	0,0427512	0,0015709	0,0199685	0,0176543	0,0378654
14	-0,0543145	0,0229542	-0,0475018	-0,0402728	-0,0190123
15	0,0502395	0,0423884	-0,0455073	-0,0187065	-0,0903322
16	0,0595230	0,1054800	0,1526687	0,1368112	0,0686641
17	-0,0674122	-0,0686692	-0,0309194	-0,0610002	0,1004308
18	-0,0876801	-0,0185949	-0,0770395	-0,0729968	-0,0959567
19	0,0013633	-0,0113226	-0,0389687	-0,0647299	0,0388989
20	-0,0885546	-0,0561932	0,1119229	0,0904823	0,1279445
21	-0,0796688	-0,0696255	-0,0779435	-0,0687524	-0,1249885
22	0,0506244	0,0928736	0,2035190	0,1719493	0,1238285
23	-0,0460508	-0,1166601	0,0577083	0,0295984	0,1397261
24	0,0415129	0,04686947	0,38544169	0,31382323	0,24948754
25	0,107798	0,00267958	0,40492251	0,36256822	0,17181691
26	-0,055164	-0,0599216	-0,1632869	-0,1443871	-0,06533062
27	-0,051725	-0,0157625	-0,1721986	-0,1485328	-0,06496708
28	-0,1187269	-0,0585772	-0,114344	-0,0891004	-0,06668505

Están resaltados en **negrita** los valores estadísticamente significativos.

ANEXO G (2)
Cuadro de correlaciones de los ítems de la encuesta

	EDAD	trabajo	Nota CBT	Aciertos	CBT -aciertos
29	0,1258981	0,0125389	0,0961821	0,0728553	0,0642757
30	0,1021493	0,0311518	0,2079822	0,1602129	0,1567073
31	-0,0304536	0,0656367	-0,0791999	-0,0581899	-0,0730316
32	-0,1066192	0,0181427	-0,1125680	-0,1121511	-0,0369522
33	0,0485223	0,0439220	-0,0286352	-0,0238598	0,0265963
34	0,1682643	0,1259350	0,1636822	0,1301414	0,1491441
35	0,0076903	-0,0271105	0,0153401	-0,0115603	0,0845960
36	-0,0750239	-0,0158105	-0,2440985	-0,2467873	-0,0342916
37	0,2094403	0,0719578	0,0885856	0,0867637	0,0298902
38	0,1173949	-0,0195172	0,2655972	0,2024896	0,2244320
39	0,0239475	-0,0119765	-0,0940671	-0,1015025	-0,0006926
40	-0,0087299	0,0734560	-0,1268663	-0,1115622	-0,0826529
41	0,0949476	-0,0931503	0,2922212	0,2457663	0,1669489
42	0,2634955	0,1108302	0,0897898	0,0674944	0,1265930
43	0,0299605	0,0004892	-0,0837893	-0,0909649	0,0146692
44	-0,1913904	-0,0452840	-0,1370782	-0,1511764	-0,0003634
45	0,1163867	-0,0030549	0,1737665	0,1225460	0,1517648
46	0,2848744	0,0322569	0,1798165	0,1327927	0,1422201
47	-0,0900508	0,0152689	-0,2663042	-0,2730502	-0,0213641
48	-0,0222546	0,0676107	-0,1031278	-0,0864581	-0,0533380
49	-0,0050395	-0,0151088	0,3343309	0,2973065	0,0958327
50	-0,0254259	-0,0082986	0,1165817	0,1048180	0,0892169
51	0,0321537	-0,0157528	0,2858566	0,2818462	0,0374318
52	-0,0271146	-0,0999000	0,2379238	0,2005250	0,1583219
53	0,1284667	-0,0519409	0,3028120	0,2635376	0,1612006
suma motivo profundo	0,1880248	0,0184826	0,1790815	0,1446278	0,1284678
suma estrategia profunda	0,2758325	0,0832743	0,2634869	0,2021937	0,2288878
suma motivo superficial	-0,0180066	0,0098419	-0,1581900	-0,1689485	0,0075738
suma estrategia superficial	-0,1282012	0,0282895	-0,2193783	-0,2160914	-0,0585387
AUTORREG	0,0205674	-0,0494810	0,3443231	0,3098289	0,1476422

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (3)

Cuadro de correlaciones de los ítems de la encuesta

	2	3	4	5	6	7
2	1					
3	-0,401661	1,000000				
4	0,049526	-0,044461	1,000000			
5	0,109673	0,019149	0,099843	1,000000		
6	0,111088	0,065854	0,107968	0,488039	1,000000	
7	0,135000	0,110094	-0,074353	0,251192	0,231471	1,000000
8	0,292391	-0,153192	-0,013205	0,369830	0,068000	0,148713
9	0,069857	0,065162	0,024464	0,144274	0,282389	0,084617
10	0,119234	0,030047	0,113532	0,195013	0,164090	0,318280
11	0,012919	0,082924	0,123362	0,328368	0,110555	0,095285
12	-0,049199	0,113482	0,179279	0,245290	0,400130	0,067064
13	0,085493	-0,002376	0,026695	0,183089	0,117726	0,239511
14	0,043622	0,010721	-0,053554	-0,173128	-0,024437	0,089574
15	0,218066	-0,179676	-0,079486	0,352143	0,111148	0,104719
16	-0,045948	0,242928	-0,161589	0,138656	0,313840	0,132830
17	0,031805	-0,008546	0,254356	0,192119	0,154128	0,141705
18	0,126433	-0,047254	0,016064	0,232810	0,106746	0,204978
19	0,012425	0,143061	0,116667	0,175815	0,322167	0,078310
20	0,153837	0,122985	0,118221	0,127786	0,201529	0,345448
21	0,244067	-0,187146	-0,173105	0,367669	0,091859	0,194550
22	0,033064	0,369765	-0,077760	0,161680	0,282759	0,202437
23	0,042527	0,052092	0,348266	0,234667	0,270567	0,196637
24	-0,130804	0,533802	-0,033316	0,011703	0,073285	0,124262
25	-0,037476	0,199115	0,151771	0,045712	0,039263	-0,031748
26	-0,042082	-0,052391	0,016195	0,172036	0,089863	0,107112
27	0,010809	-0,164545	-0,009819	0,135748	0,177153	0,065533
28	-0,058767	-0,039369	-0,145126	-0,012757	0,043175	0,184256

Están resaltados en **negrita** los valores estadísticamente significativos.

ANEXO G (4)
Cuadro de correlaciones de los ítems de la encuesta

	2	3	4	5	6	7
29	-0,016855	0,239138	0,043906	0,131908	0,230700	0,177599
30	0,020513	0,183333	-0,091580	0,120026	0,072120	0,037749
31	-0,129601	0,030874	-0,095786	-0,048263	-0,252472	-0,058289
32	-0,111144	0,028900	-0,127221	-0,039033	-0,112399	0,002862
33	-0,077463	0,040452	0,029480	0,131061	0,008471	0,074402
34	-0,014719	0,193150	0,097547	0,073964	0,154546	0,071058
35	-0,021199	0,017403	-0,047882	-0,089435	-0,125548	-0,071464
36	0,050271	-0,138585	0,187478	0,053967	-0,034917	0,057617
37	-0,008627	0,120095	0,102457	0,126449	0,157074	0,033212
38	0,051967	0,127490	0,039975	0,132136	0,122072	0,018509
39	0,056537	-0,048651	0,232832	0,038936	-0,074115	-0,100895
40	-0,054188	-0,136116	-0,051469	-0,042878	-0,167884	-0,028668
41	0,000575	0,129633	0,078780	0,189322	0,170454	-0,042569
42	0,010329	0,161074	0,052647	0,052776	0,036741	-0,114306
43	0,005917	-0,082193	0,087417	-0,013672	-0,067491	-0,031998
44	0,094001	-0,208571	-0,064402	-0,051905	-0,045053	0,036598
45	0,039655	0,102876	0,027457	0,085035	0,069622	0,020399
46	0,045144	0,102749	0,055367	0,225178	0,183484	-0,043067
47	0,001457	-0,164413	-0,018364	0,024942	0,050090	0,049480
48	0,058739	-0,101168	0,028044	0,061947	-0,024438	0,060849
49	0,073324	0,154413	-0,054216	0,129295	0,152660	0,179432
50	0,056453	0,035878	-0,034442	-0,044247	-0,071815	0,034071
51	0,049366	0,140628	-0,135238	0,084354	0,029698	0,140289
52	0,094651	0,054959	-0,038135	0,154642	0,086300	0,152640
53	0,116861	0,112607	-0,087526	0,108400	0,192112	0,094980
suma MP	-0,018009	0,190323	0,086107	0,196970	0,189706	0,083057
suma EP	0,044401	0,214725	0,052244	0,174722	0,163313	-0,011551
suma MS	-0,024386	-0,075776	0,054097	-0,030428	-0,151883	-0,073673
suma ES	0,014750	-0,168402	-0,008816	-0,005585	-0,113205	0,042329
AUTORREG.	0,105975	0,133990	-0,096021	0,106389	0,087606	0,161183

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (5)

Cuadro de correlaciones de los ítems de la encuesta

	8	9	10	11	12	13
8	1					
9	0,465941	1,000000				
10	0,007830	-0,006055	1,000000			
11	0,195470	0,094056	0,250702	1,000000		
12	-0,051768	0,207885	0,280369	0,590668	1,000000	
13	0,217496	0,245837	-0,000153	-0,283113	-0,054814	1,000000
14	0,002772	0,069565	0,008694	-0,103595	-0,020393	0,087933
15	0,450729	0,131812	-0,017728	0,109764	0,009551	0,151999
16	0,123653	0,322022	-0,014499	0,081131	0,226379	0,101441
17	-0,082820	-0,076808	0,561698	0,236549	0,308785	-0,026185
18	0,384722	0,158815	-0,002794	0,174009	0,041472	0,191161
19	0,086888	0,269670	0,048207	0,082661	0,213454	0,110410
20	-0,022240	0,031166	0,527618	0,168594	0,235093	0,076336
21	0,513526	0,198429	-0,043905	0,162565	-0,056920	0,211213
22	0,176764	0,369233	0,069132	0,075320	0,148421	0,223745
23	-0,073110	0,016871	0,562463	0,269117	0,392737	-0,019786
24	-0,033646	0,006047	0,033780	0,006817	-0,003856	0,021829
25	-0,017210	0,052569	0,087139	0,019313	0,053603	-0,003170
26	0,145597	0,113427	-0,051258	0,052080	0,028791	0,144971
27	0,116375	0,159430	-0,053805	0,036174	0,143274	0,145258
28	0,000986	0,052149	-0,015642	-0,010560	0,040488	0,141084

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (6)

Cuadro de correlaciones de los ítems de la encuesta

	8	9	10	11	12	13
29	0,079103	0,246168	0,158311	0,092061	0,120746	0,107252
30	0,102191	0,178602	0,089370	0,021923	-0,014620	0,128465
31	-0,140035	-0,223938	-0,124984	0,067199	-0,056915	-0,175610
32	-0,012881	-0,050304	-0,161536	-0,033079	0,004686	0,014433
33	0,096379	0,146103	0,065749	0,139556	0,059171	0,059609
34	0,013846	0,170790	0,136368	0,009015	0,108363	0,139258
35	-0,151495	-0,109453	-0,121763	-0,050884	-0,104271	-0,031937
36	-0,069324	-0,131886	-0,078013	0,074565	0,046000	0,002331
37	0,088559	0,147704	0,120446	0,128655	0,105896	0,039607
38	0,161721	0,226516	0,095624	0,081523	0,074450	0,058415
39	-0,081950	-0,085719	-0,045121	0,088840	0,110898	0,053906
40	-0,018202	-0,025443	-0,132018	-0,022466	-0,088773	-0,030964
41	0,151763	0,303725	-0,009944	0,042944	0,051149	0,100275
42	-0,013232	0,004364	0,045261	0,039544	0,021342	-0,068849
43	-0,106294	-0,202243	-0,073518	-0,009879	0,046596	-0,032016
44	-0,012258	-0,122576	-0,042649	0,018663	-0,038476	-0,028096
45	0,041959	0,223590	0,088236	-0,007973	-0,033486	0,123434
46	0,135054	0,211488	0,092158	-0,017834	0,004443	0,128257
47	-0,017159	-0,115934	-0,030235	0,015150	0,008536	0,013556
48	-0,031922	-0,062076	-0,078661	-0,039619	-0,060422	0,020540
49	0,169241	0,208301	0,083208	0,067311	0,044301	0,135127
50	0,135228	0,121075	0,032121	0,021655	-0,039720	0,041728
51	0,173775	0,140644	0,131313	-0,004405	-0,084708	0,178875
52	0,238399	0,180419	0,141983	0,068169	0,013531	0,089612
53	0,218517	0,313311	0,110025	0,062207	0,058515	0,134391
suma MP	0,134047	0,316401	0,136464	0,122401	0,091614	0,129025
suma EP	0,112916	0,229045	0,136268	0,036300	0,053245	0,113536
suma MS	-0,163083	-0,230451	-0,129274	0,034431	-0,002017	-0,049593
suma ES	-0,043985	-0,120397	-0,144298	0,000332	-0,038486	-0,005606
AUTORREG.	0,257208	0,258972	0,133792	0,056327	-0,010971	0,156438

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (7)

Cuadro de correlaciones de los ítems de la encuesta

	15	16	17	18	19	20
15	1					
16	0,214036	1				
17	-0,119622	-0,149302	1,000000			
18	0,463206	0,205787	-0,025124	1		
19	0,112600	0,345874	0,047660	0,487070	1	
20	-0,067837	0,038398	0,479972	-0,102705	-0,041871	1,000000
21	0,557990	0,148209	-0,149722	0,508762	0,140037	-0,069416
22	0,079044	0,604517	-0,079402	0,229058	0,440944	0,134797
23	-0,079792	-0,019281	0,684715	0,015759	0,127401	0,591062
24	-0,099179	0,149056	-0,058732	-0,008391	0,066945	0,129327
25	-0,066183	-0,007592	0,058079	-0,087518	0,046441	0,108307
26	0,146428	0,004047	-0,054057	0,235969	0,107624	-0,102090
27	0,080565	0,055449	-0,018168	0,148919	0,162517	-0,070807
28	0,018484	0,030287	0,027018	0,108757	0,015842	0,057328

Están resaltados en **negrita** los valores estadísticamente significativos

	21	22	23	26	27	28
21	1					
22	0,261465	1				
23	-0,106553	0,079471	1,000000			
24	-0,072315	0,297646	0,034365	-0,147200	-0,190315	-0,051880
25	-0,078338	0,072629	0,107531	-0,114054	-0,089445	-0,102095
26	0,261933	0,034907	-0,022875	1		
27	0,163655	0,024387	0,003757	0,809704	1,000000	
28	0,137162	0,036154	0,024977	0,608693	0,704304	1

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (8)
Cuadro de correlaciones de los ítems de la encuesta

	15	16	17	18	19	20
29	0,081255	0,230728	0,118438	0,059949	0,240612	0,196066
30	-0,044215	0,164154	-0,044872	-0,108548	0,081747	0,031326
31	-0,009171	-0,201908	0,006790	-0,103148	-0,224918	-0,044258
32	0,144934	0,041075	-0,039594	0,111804	0,056301	-0,069682
33	0,020254	0,016580	0,052486	0,070343	0,167858	0,022272
34	-0,079045	0,138017	0,080344	0,007818	0,142233	0,144716
35	-0,043128	-0,025224	-0,106384	-0,116821	-0,079938	-0,107850
36	0,041149	-0,148216	0,088795	0,064807	-0,024340	-0,024984
37	-0,006504	0,080099	0,059075	-0,029710	0,153197	0,098188
38	-0,130223	0,087621	0,037185	-0,054722	0,062658	0,072894
39	-0,008906	-0,151786	0,154022	-0,143594	-0,108998	0,038488
40	0,068606	0,014354	-0,026467	-0,049426	-0,155523	-0,048905
41	-0,005859	0,051099	0,000383	-0,095124	0,053962	0,035166
42	-0,092487	-0,037969	-0,038852	-0,149239	0,051289	0,013838
43	0,061986	-0,063658	0,123292	0,024373	-0,077237	-0,079722
44	0,026053	-0,027461	0,030507	-0,009813	-0,118658	-0,008932
45	-0,052004	0,099066	-0,024425	-0,040968	0,214251	-0,059204
46	0,079504	0,057002	0,047911	0,033676	0,114338	-0,009978
47	0,022636	-0,101950	0,009186	-0,006066	-0,059140	-0,044801
48	0,137379	-0,045977	-0,033376	0,046629	-0,045560	-0,069054
49	-0,023496	0,143975	-0,025205	0,049260	0,065315	0,124600
50	-0,023721	0,003726	-0,099301	0,070566	0,078086	0,155482
51	0,004351	0,045156	0,004674	-0,031366	0,017809	0,087297
52	0,038182	0,105870	-0,011670	0,034595	0,050799	0,118518
53	-0,064683	0,103471	-0,021420	-0,005172	0,102704	0,078898
suma MP	0,008683	0,146658	0,063400	-0,008081	0,260849	0,087217
suma EP	-0,080544	0,118335	0,031509	-0,079133	0,127821	0,072142
suma MS	0,001397	-0,173338	0,052604	-0,122131	-0,175967	-0,076305
suma ES	0,126534	-0,050069	0,007869	0,051749	-0,082979	-0,069357
AUTORREG.	-0,018487	0,102842	-0,050239	0,037957	0,088178	0,164774

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (9)

Cuadro de correlaciones de los ítems de la encuesta

	21	22	23	26	27	28
29	0,081550	0,338281	0,209236	0,010898	0,069709	0,047644
30	-0,004546	0,290520	0,013985	-0,024076	-0,036244	-0,095815
31	-0,081092	-0,239784	-0,083451	-0,091634	-0,140164	-0,048008
32	0,124148	0,036055	-0,109552	0,034323	0,022953	0,027541
33	0,137800	0,182935	0,095921	0,147739	0,144172	0,159940
34	-0,024478	0,284324	0,139333	-0,007297	0,030817	0,007482
35	-0,089052	-0,106211	-0,122468	-0,182241	-0,220698	-0,180248
36	0,070931	-0,183203	0,106788	0,258591	0,217530	0,175212
37	0,033963	0,196045	0,182340	0,016782	-0,003672	0,008402
38	-0,008179	0,268351	0,108774	0,004097	-0,013847	-0,107595
39	-0,078084	-0,169090	0,184465	-0,062202	-0,039495	-0,069251
40	0,032653	-0,128098	-0,063257	0,014449	0,013517	0,037228
41	0,061228	0,209947	0,065191	0,114134	0,118138	0,019100
42	-0,022363	0,118727	0,024396	-0,045030	-0,022951	-0,084554
43	-0,032012	-0,117967	0,040893	0,018944	0,040089	0,093122
44	0,074231	-0,144479	-0,092556	-0,030819	0,028410	0,088570
45	0,050646	0,180088	0,033838	0,013102	0,047488	-0,026205
46	0,070993	0,156400	0,096676	0,043046	0,098853	0,000840
47	0,053680	-0,132185	-0,044879	0,017280	0,048542	0,011561
48	0,073579	-0,125320	-0,082413	0,139289	0,166702	0,197266
49	0,014464	0,239037	0,085128	-0,124856	-0,090617	-0,117386
50	0,027473	0,116194	0,014165	-0,129385	-0,104785	-0,087596
51	0,114829	0,268964	0,064369	-0,040489	-0,045648	-0,100436
52	0,017384	0,180168	0,081323	-0,024332	-0,057379	-0,021325
53	0,047006	0,252806	0,098344	-0,046609	-0,008331	-0,044839
suma MP	0,109450	0,334725	0,181847	0,085646	0,107071	0,061237
suma EP	0,002199	0,319794	0,115479	-0,011592	0,015184	-0,083079
suma MS	-0,077702	-0,246043	-0,008991	-0,108700	-0,114141	-0,079381
suma ES	0,114558	-0,166870	-0,075356	0,120245	0,133292	0,158591
AUTORREG.	0,062551	0,288004	0,089692	-0,109397	-0,092722	-0,107895

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (10)

Cuadro de correlaciones de los ítems de la encuesta

	29	30	31	32	33	34
29	1					
30	0,287098	1,000000				
31	-0,143477	-0,173824	1			
32	0,054896	-0,117501	0,233311	1,000000		
33	0,230960	0,099892	-0,049979	-0,045014	1	
34	0,318519	0,279334	-0,168898	-0,214520	0,341145	1
35	-0,142706	-0,029074	0,316429	0,022539	-0,060564	-0,144296
36	-0,104581	-0,305306	0,247865	0,143358	0,080301	-0,057313
37	0,308869	0,176372	-0,130147	-0,192185	0,375236	0,377825
38	0,222438	0,319897	-0,159520	-0,185812	0,084893	0,332880
39	-0,048310	-0,135875	0,234394	0,062251	0,051204	-0,082055
40	-0,067523	-0,123485	0,289582	0,438217	-0,008110	-0,168981
41	0,335219	0,292598	-0,284593	-0,188476	0,301287	0,402825
42	0,163569	0,242891	-0,088557	-0,292543	0,177180	0,487433
43	-0,125671	-0,124915	0,222343	0,192440	-0,090803	-0,046460
44	-0,038501	-0,111567	0,176236	0,306022	-0,009978	-0,185957
45	0,184439	0,306930	-0,169629	-0,203324	0,224602	0,258118
46	0,248862	0,213058	-0,143646	-0,188722	0,217475	0,380979
47	-0,141952	-0,095271	0,165084	0,176081	-0,101722	-0,109917
48	-0,008910	-0,113850	0,282765	0,203005	0,090222	-0,058677
49	0,330678	0,320202	-0,142557	-0,043739	0,089008	0,239050
50	0,101113	0,116135	-0,120319	-0,208845	0,095192	0,135542
51	0,258522	0,336433	-0,174222	-0,105220	0,198054	0,306038
52	0,189538	0,138617	-0,220071	-0,127127	0,153805	0,150404
53	0,189688	0,346688	-0,305385	-0,143942	0,137887	0,239692
suma MP	0,613950	0,350057	-0,228244	-0,179249	0,652112	0,513707
suma EP	0,355457	0,583727	-0,210110	-0,295815	0,271470	0,718066
suma MS	-0,193349	-0,175173	0,614406	0,207794	-0,076386	-0,182156
suma ES	-0,048300	-0,233192	0,371953	0,636972	0,030896	-0,211209
AUTORREG.	0,290744	0,338986	-0,260043	-0,186492	0,187812	0,295149

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (11)

Cuadro de correlaciones de los ítems de la encuesta

	35	36	37	38	39	40
35	1					
36	0,130298	1				
37	-0,061634	-0,082861	1			
38	0,042979	-0,260785	0,254646	1		
39	0,165639	0,536141	0,025348	-0,127468	1	
40	0,149942	0,271361	-0,230971	-0,223872	0,247159	1
41	-0,151595	-0,127778	0,380186	0,354285	-0,028649	-0,146867
42	0,025592	-0,092292	0,391276	0,297019	-0,015448	-0,244920
43	0,147643	0,281847	-0,178964	-0,193159	0,246567	0,190048
44	0,199146	0,248242	-0,219667	-0,131551	0,121729	0,330245
45	0,056115	-0,066657	0,225110	0,253535	-0,100693	-0,126563
46	-0,057427	-0,018040	0,341756	0,323747	0,012354	-0,206133
47	0,262483	0,251358	-0,189373	-0,151837	0,145670	0,370585
48	0,183386	0,282843	-0,087743	-0,134501	0,215072	0,276212
49	-0,149823	-0,184298	0,050823	0,319366	-0,080973	-0,051263
50	-0,064926	-0,153453	0,129730	0,195726	-0,075759	-0,215444
51	-0,210992	-0,196665	0,173208	0,355739	-0,119184	-0,120947
52	-0,188284	-0,127647	0,155128	0,263278	-0,090681	-0,252163
53	-0,129712	-0,325629	0,219348	0,489030	-0,162508	-0,210464
suma MP	-0,099068	-0,088408	0,725092	0,350298	-0,030398	-0,181708
suma EP	-0,046859	-0,208661	0,449964	0,672965	-0,098580	-0,277865
suma MS	0,659101	0,470738	-0,161356	-0,176650	0,603152	0,400488
suma ES	0,211091	0,581045	-0,246372	-0,281754	0,351542	0,685088
AUTORREG.	-0,202931	-0,270091	0,202926	0,439627	-0,144807	-0,243542

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (12)

Cuadro de correlaciones de los ítems de la encuesta

	41	42	43	44
41	1			
42	0,359142	1		
43	-0,114903	-0,106610	1	
44	-0,122594	-0,284190	0,327367	1
45	0,306776	0,382183	-0,181774	-0,087529
46	0,343950	0,456133	-0,138518	-0,243607
47	-0,227351	-0,109090	0,328495	0,483394
48	-0,108622	-0,056624	0,262370	0,387907
49	0,262652	0,028701	-0,132119	-0,007246
50	0,069986	0,169095	-0,059909	-0,100034
51	0,297380	0,206575	-0,214825	-0,057743
52	0,323687	0,150096	-0,129005	-0,133923
53	0,372811	0,215915	-0,186743	-0,122491
suma MP	0,665823	0,454836	-0,216477	-0,150788
suma EP	0,509153	0,728654	-0,174403	-0,277120
suma MS	-0,255882	-0,085289	0,583527	0,414015
suma ES	-0,211987	-0,297997	0,387185	0,714276
AUTORREGULACIÓN	0,350173	0,219558	-0,195202	-0,119486

Están resaltados en **negrita** los valores estadísticamente significativos

	45	46	47	48
45	1			
46	0,37418	1		
47	-0,054775	-0,149645	1	
48	0,022903	0,062079	0,329075	1
49	0,170457	0,156058	-0,061831	-0,062203
50	0,123318	0,051111	-0,100498	-0,067762
51	0,310250	0,234957	-0,003013	-0,096588
52	0,197483	0,208470	-0,222572	-0,112796
53	0,287237	0,330814	-0,173722	-0,107366
suma MP	0,612751	0,468160	-0,212095	-0,024348
suma EP	0,464541	0,705653	-0,176158	-0,080834
suma MS	-0,131149	-0,144945	0,607084	0,406042
suma ES	-0,139261	-0,184788	0,494864	0,661066
AUTORREGULACIÓN	0,297482	0,257857	-0,152826	-0,123572

Están resaltados en **negrita** los valores estadísticamente significativos

ANEXO G (13)

Cuadro de correlaciones de los ítems de la encuesta

	49	50	51	52	53
49	1				
50	0,2180414	1			
51	0,5475130	0,3454217	1		
52	0,3241034	0,3940654	0,3760990	1	
53	0,4370758	0,2583944	0,4207463	0,496305	1
suma MP	0,261786	0,162762	0,372700	0,300843	0,360800
suma EP	0,308124	0,190361	0,419981	0,266911	0,474809
suma MS	-0,183799	-0,135694	-0,233850	-0,275197	-0,303837
suma ES	-0,106143	-0,225121	-0,173739	-0,225375	-0,271362
AUTORREGULACIÓN.	0,681911	0,693965	0,754218	0,716129	0,690933

Están resaltados en **negrita** los valores estadísticamente significativos

	suma MP	suma EP	suma MS	suma ES	AUTORREG.
suma MP	1				
suma EP	0,623893	1			
suma MS	-0,242075	-0,218381	1		
suma ES	-0,189464	-0,348454	0,578078	1	
AUTORREGULACIÓN	0,397775	0,451021	-0,307699	-0,283482	1

Están resaltados en **negrita** los valores estadísticamente significativos.

ANEXO H (1)

Análisis factorial con el SPSS de los 20 ítems del RSPQ-2F, con 5 componentes

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,811
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1345,304
	gl	190
	Sig.	,000

Comunalidades

	Inicial	Extracción
p29	1,000	,538
p30	1,000	,480
p31	1,000	,642
p32	1,000	,634
p33	1,000	,464
p34	1,000	,539
p35	1,000	,674
p36	1,000	,676
p37	1,000	,503
p38	1,000	,464
p39	1,000	,564
p40	1,000	,542
p41	1,000	,521
p42	1,000	,593
p43	1,000	,468
p44	1,000	,649
p45	1,000	,423
p46	1,000	,483
p47	1,000	,635
p48	1,000	,443

Método de extracción: Análisis de Componentes principales.

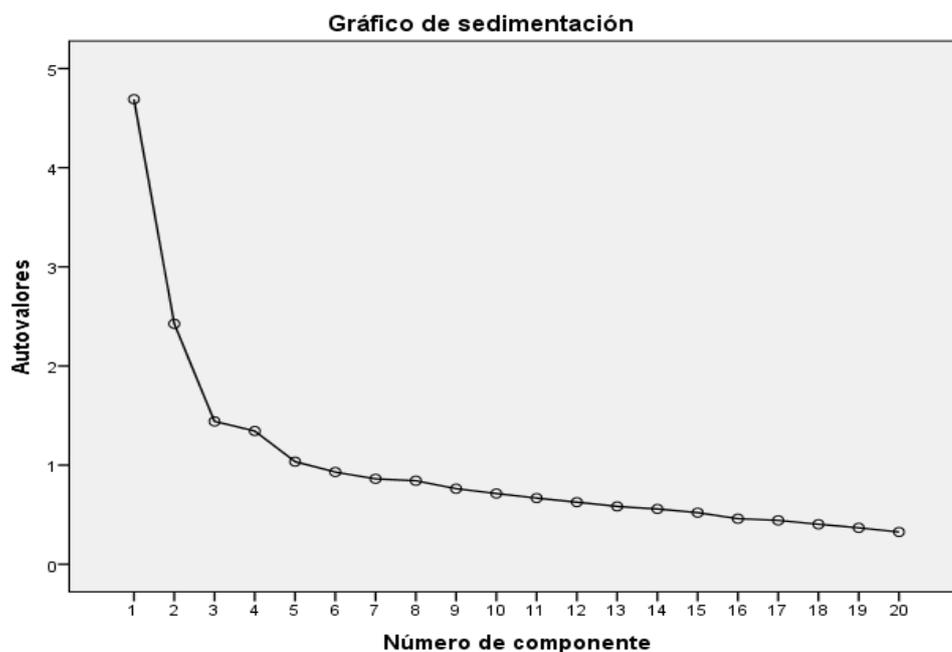
ANEXO H (2)

Análisis factorial con el SPSS de los 20 ítems del RSPQ-2F, con 5 componentes

Varianza total explicada					
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	4,690	23,451	23,451	4,690	23,451
2	2,425	12,126	35,577	2,425	12,126
3	1,441	7,204	42,780	1,441	7,204
4	1,344	6,720	49,500	1,344	6,720
5	1,035	5,175	54,675	1,035	5,175
6	,930	4,650	59,325		
7	,862	4,309	63,634		
8	,842	4,212	67,846		
9	,762	3,811	71,658		
10	,713	3,567	75,225		
11	,668	3,338	78,563		
12	,626	3,132	81,696		
13	,583	2,916	84,611		
14	,558	2,788	87,399		
15	,521	2,603	90,002		
16	,460	2,300	92,302		
17	,443	2,213	94,515		
18	,404	2,018	96,533		
19	,368	1,838	98,372		
20	,326	1,628	100,000		

Varianza total explicada				
Componente	Sumas de las saturaciones al cuadrado de la extracción	Suma de las saturaciones al cuadrado de la rotación		
	% acumulado	Total	% de la varianza	% acumulado
1	23,451	3,658	18,288	18,288
2	35,577	2,175	10,873	29,161
3	42,780	1,994	9,970	39,131
4	49,500	1,650	8,249	47,380
5	54,675	1,459	7,295	54,675

ANEXO H (3)



Matriz de componentes ^a					
	Componente				
	1	2	3	4	5
p41	,612				
p42	,604	,327			
p34	,599	,382			
p37	,575	,306			
p46	,563	,371			
p38	,546			,307	
p40	-,531	,331			
p44	-,521	,363	,422		
p45	,491	,307			
p47	-,486	,376			-,331
p30	,482		,436		
p32	-,470		,427	-,326	,317
p43	-,440	,332			-,401
p29	,422		,371	-,341	
p48	-,356	,550			
p36	-,427	,500	-,434		
p39		,481	-,470		
p33	,330	,440		-,391	
p35				,634	,331
p31	-,461				,549

Método de extracción: Análisis de componentes principales. a. 5 componentes extraídos

ANEXO H (4)

Matriz de componentes rotados^a					
	Componente				
	1	2	3	4	5
p34	,700				
p46	,658				
p41	,658				
p42	,646			-,372	
p37	,641				
p33	,563				
p45	,554				
p29	,533			,412	
p38	,472		-,437		
p47		,769			
p44		,730		,309	
p43		,605			
p48		,507			
p36			,758		
p39			,698		
p30	,422		-,547		
p32				,747	
p40		,326		,590	
p35					,784
p31				,307	,657

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.^a

a. La rotación ha convergido en 11 iteraciones.

Matriz de transformación de las componentes

Componente	1	2	3	4	5
1	,739	-,436	-,362	-,295	-,212
2	,671	,512	,383	,250	,281
3	,017	,366	-,744	,533	-,167
4	-,059	,284	-,407	-,550	,669
5	,018	-,577	-,058	,513	,633

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

ANEXO H (5)
Análisis factorial de los ítems del RSPQ-2F
PARA 4 COMPONENTES

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,811
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1345,304
	gl	190
	Sig.	,000

	Inicial	Extracción
p29	1,000	,514
p30	1,000	,475
p31	1,000	,341
p32	1,000	,534
p33	1,000	,455
p34	1,000	,509
p35	1,000	,564
p36	1,000	,660
p37	1,000	,483
p38	1,000	,449
p39	1,000	,562
p40	1,000	,494
p41	1,000	,502
p42	1,000	,592
p43	1,000	,307
p44	1,000	,585
p45	1,000	,423
p46	1,000	,483
p47	1,000	,526
p48	1,000	,442

ANEXO H (6)

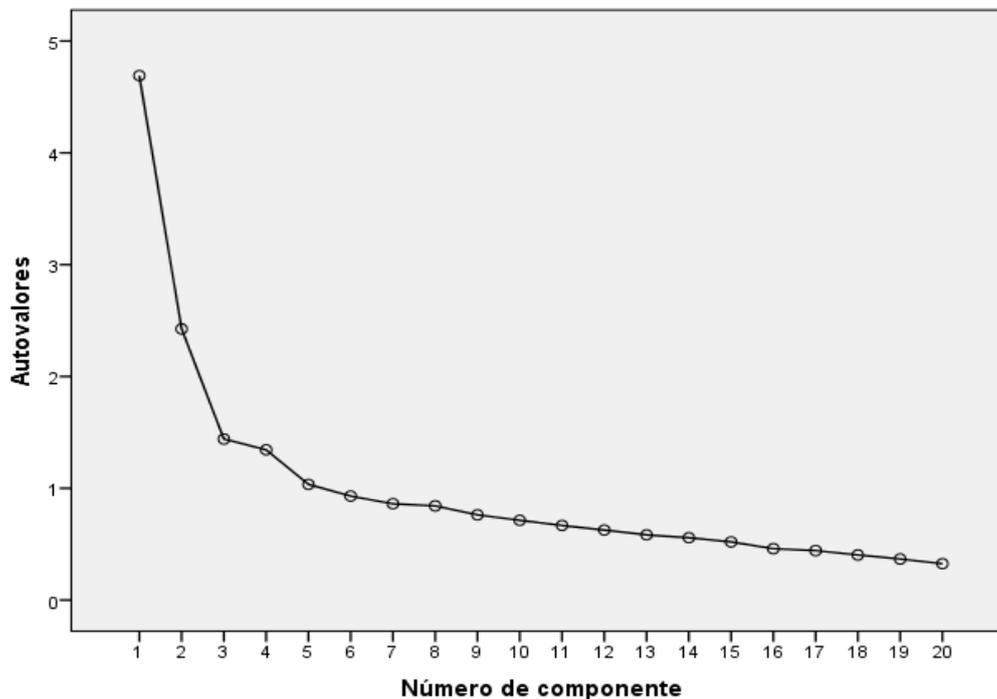
Análisis factorial de los ítems del RSPQ-2F para 4 COMPONENTES

Varianza total explicada					
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	4,690	23,451	23,451	4,690	23,451
2	2,425	12,126	35,577	2,425	12,126
3	1,441	7,204	42,780	1,441	7,204
4	1,344	6,720	49,500	1,344	6,720
5	1,035	5,175	54,675		
6	,930	4,650	59,325		
7	,862	4,309	63,634		
8	,842	4,212	67,846		
9	,762	3,811	71,658		
10	,713	3,567	75,225		
11	,668	3,338	78,563		
12	,626	3,132	81,696		
13	,583	2,916	84,611		
14	,558	2,788	87,399		
15	,521	2,603	90,002		
16	,460	2,300	92,302		
17	,443	2,213	94,515		
18	,404	2,018	96,533		
19	,368	1,838	98,372		
20	,326	1,628	100,000		

Componente	Sumas de las saturaciones al cuadrado de la extracción	Suma de las saturaciones al cuadrado de la rotación		
	% acumulado	Total	% de la varianza	% acumulado
1	23,451	3,660	18,299	18,299
2	35,577	2,700	13,498	31,796
3	42,780	2,066	10,328	42,124
4	49,500	1,475	7,376	49,500

ANEXO H (7)

Análisis factorial de los ítems del RSPQ-2F para 4 COMPONENTES

Gráfico de sedimentación

Matriz de componentes^a				
	Componente			
	1	2	3	4
p41	,612			
p42	,604	,327		
p34	,599	,382		
p37	,575	,306		
p46	,563	,371		
p38	,546			,307
p40	-,531	,331		
p44	-,521	,363	,422	
p45	,491	,307		
p47	-,486	,376		
p30	,482		,436	
p32	-,470		,427	-,326
p31	-,461			
p43	-,440	,332		
p29	,422		,371	-,341
p48	-,356	,550		
p36	-,427	,500	-,434	
p39		,481	-,470	
p33	,330	,440		-,391
p35				,634

ANEXO H (8)

Análisis factorial de los ítems del RSPQ-2F para 4 COMPONENTES

Matriz de componentes rotados^a				
	Componente			
	1	2	3	4
34	,695			
42	,680			
46	,669			
41	,639			
37	,618			
45	,587			
33	,514			-,375
38	,505		-,397	
29	,486			-,451
44		,746		
40		,647		
47		,618		,361
32		,613		-,301
48		,569		
43		,417	,302	
36			,769	
39			,724	
30	,442		-,524	
31			,369	,307
35				,701

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.^a

a. La rotación ha convergido en 12 iteraciones.

Matriz de transformación de las componentes				
Componente	1	2	3	4
1	,740	-,533	-,368	-,183
2	,670	,583	,444	,117
3	,009	,612	-,758	-,224
4	,061	-,030	-,304	,950

ANEXO H (9)

Análisis factorial de los ítems del RSPQ-2F para 2 COMPONENTES

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,811
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1345,304
	gl	190
	Sig.	,000

Comunalidades

	Inicial	Extracción
p29	1,000	,260
p30	1,000	,243
p31	1,000	,286
p32	1,000	,245
p33	1,000	,302
p34	1,000	,505
p35	1,000	,157
p36	1,000	,432
p37	1,000	,425
p38	1,000	,312
p39	1,000	,320
p40	1,000	,392
p41	1,000	,460
p42	1,000	,472
p43	1,000	,304
p44	1,000	,404
p45	1,000	,335
p46	1,000	,455
p47	1,000	,378
p48	1,000	,429

Método de extracción:
Análisis de Componentes principales

ANEXO H (10)**Análisis factorial de los ítems del RSPQ-2F para 2 COMPONENTES**

Varianza total explicada					
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	4,690	23,451	23,451	4,690	23,451
2	2,425	12,126	35,577	2,425	12,126
3	1,441	7,204	42,780		
4	1,344	6,720	49,500		
5	1,035	5,175	54,675		
6	,930	4,650	59,325		
7	,862	4,309	63,634		
8	,842	4,212	67,846		
9	,762	3,811	71,658		
10	,713	3,567	75,225		
11	,668	3,338	78,563		
12	,626	3,132	81,696		
13	,583	2,916	84,611		
14	,558	2,788	87,399		
15	,521	2,603	90,002		
16	,460	2,300	92,302		
17	,443	2,213	94,515		
18	,404	2,018	96,533		
19	,368	1,838	98,372		
20	,326	1,628	100,000		

Varianza total explicada				
Componente	Sumas de las saturaciones al cuadrado de la extracción	Suma de las saturaciones al cuadrado de la rotación		
	% acumulado	Total	% de la varianza	% acumulado
1	23,451	3,798	18,989	18,989
2	35,577	3,317	16,587	35,577

ANEXO H (11)

Análisis factorial de los ítems del RSPQ-2F para 2 COMPONENTES

Matriz de componentes^a		
	Componente	
	1	2
p41	,612	
p42	,604	,327
p34	,599	,382
p37	,575	,306
p46	,563	,371
p38	,546	
p40	-,531	,331
p44	-,521	,363
p45	,491	,307
p47	-,486	,376
p30	,482	
p32	-,470	
p31	-,461	
p43	-,440	,332
p29	,422	
p48	-,356	,550
p36	-,427	,500
p39		,481
p33	,330	,440
p35		

Método de extracción: Análisis de componentes
principales (2 componentes extraídos)

Matriz de componentes rotados^a			
	Componente		
	1	2	
p34	,706		<i>Enfoque profundo</i>
p42	,676		<i>Enfoque profundo</i>
p46	,672		<i>Enfoque profundo</i>
p41	,660		<i>Enfoque profundo</i>
p37	,640		<i>Enfoque profundo</i>
p45	,575		<i>Enfoque profundo</i>
p33	,533		<i>Enfoque profundo</i>
p29	,508		<i>Enfoque profundo</i>
p38	,500		<i>Enfoque profundo</i>

p30	,440		<i>Enfoque profundo</i>
p36		,657	Enfoque superficial
p48		,651	Enfoque superficial
p44		,610	Enfoque superficial
p47		,598	Enfoque superficial
p40		,591	Enfoque superficial
p39		,561	Enfoque superficial
p43		,534	Enfoque superficial
p31		,501	Enfoque superficial
p32		,416	Enfoque superficial
p35		,396	Enfoque superficial

Método de extracción: Análisis de componentes principales.

Método de rotación:

Normalización Varimax con Kaiser.^a

a. La rotación ha convergido en 3 iteraciones.

Matriz de transformación de las componentes		
Componente	1	2
1	,778	-,628
2	,628	,778

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

ANEXO H (12)Análisis de fiabilidad de los enfoques de aprendizaje
RSPQ-2F**Análisis de fiabilidad de los ENFOQUES PROFUNDOS****Resumen del procesamiento de los casos**

		N	%
Casos	Válidos	271	94,1
	Excluidos ^a	17	5,9
	Total	288	100,0

ENFOQUES PROFUNDOS: ítems 29, 30, 33, 34, 37, 38, 41, 42, 45 y 46

Estadísticos de fiabilidad**ENFOQUES PROFUNDOS**

Alfa de Cronbach	N de elementos
,804	10

Estadísticos total-elemento				
	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
29	34,00	42,385	,415	,794
33	34,04	42,673	,370	,799
37	35,08	38,971	,517	,783
41	34,18	42,102	,569	,780
45	35,47	40,509	,454	,790
30	34,04	43,246	,395	,795
34	35,25	39,942	,592	,774
38	34,29	41,681	,435	,792
42	35,88	40,742	,550	,779
46	35,68	40,254	,535	,780

ANEXO H (13)**Análisis de fiabilidad de los enfoques de aprendizaje (RSPQ-2F)****Análisis de la fiabilidad de los ítems de MOTIVO PROFUNDO: 29, 33, 37, 41, 45**

		N	%
Casos	Válidos	272	94,4
	Excluidos ^a	16	5,6
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,657	5

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
29	15,85	11,078	,377	,620
33	15,90	10,609	,417	,602
37	16,94	9,328	,474	,574
41	16,03	11,279	,497	,581
45	17,34	10,675	,329	,648

Análisis de la fiabilidad de los ítems de ESTRATEGIA PROFUNDA: 30, 34, 38, 42, 46

		N	%
Casos	Válidos	271	94,1
	Excluidos ^a	17	5,9
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,717	5

Estadísticos total-elemento

Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
13,52	11,643	,361	,710
14,73	10,086	,542	,642
13,77	10,471	,445	,682
15,37	10,240	,540	,643
15,17	10,161	,490	,663

ANEXO H (14)

Análisis de fiabilidad de los enfoques de aprendizaje (RSPQ-2F)

Análisis de fiabilidad de los ENFOQUES SUPERFICIALES

Ítems de ENFOQUES SUPERFICIALES: 31, 32, 35, 36, 39, 40, 43, 44, 47, 48.

Resumen del procesamiento de los casos			
		N	%
Casos	Válidos	271	94,1
	Excluidos ^a	17	5,9
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,760	10

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
32	21,86	40,333	,335	,753
36	23,21	38,902	,466	,734
40	23,04	39,213	,504	,730
44	22,59	36,917	,504	,728
48	22,99	38,448	,471	,733
31	23,65	41,057	,419	,741
35	22,44	40,958	,290	,760
39	23,16	40,238	,372	,747
43	23,61	41,705	,424	,742
47	23,33	39,586	,494	,731

ANEXO H (15)

Análisis de fiabilidad de los enfoques de aprendizaje (RSPQ-2F)

MOTIVOS SUPERFICIALES: Ítems 31, 35, 39, 43 y 47

Resumen del procesamiento de los casos			
		N	%
Casos	Válidos	272	94,4
	Excluidos ^a	16	5,6
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,580	5

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
31	9,62	8,975	,373	,509
35	8,40	7,968	,345	,525
39	9,12	8,646	,294	,553
47	9,30	8,881	,343	,523
43	9,58	9,418	,362	,519

ESTRATEGIAS SUPERFICIALES: Ítems 32, 36, 40, 44 y 48

Resumen del procesamiento de los casos			
		N	%
Casos	Válidos	271	94,1
	Excluidos ^a	17	5,9
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,668	5

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
32	10,34	12,656	,391	,631
36	11,69	13,458	,337	,653
40	11,52	12,777	,493	,590
44	11,07	11,462	,474	,592
48	11,47	12,509	,425	,616

ANEXO H (16)

Análisis de fiabilidad de los ítems de AUTORREGULACIÓN

AUTORREGULACIÓN: Ítems 49, 50, 51, 52 y 53

Resumen del procesamiento de los casos			
		N	%
Casos	Válidos	272	94,4
	Excluidos ^a	16	5,6
	Total	288	100,0

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,735	5

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
49	17,97	10,940	,503	,689
50	19,15	9,469	,404	,751
51	18,43	9,899	,578	,657
52	18,01	10,590	,545	,673
53	17,72	11,226	,537	,682

ANEXO H (17)

Análisis factorial de los ítems de la encuesta sobre tipos de examen
(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27 y 28)

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,700
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2934,064
	gl	351
	Sig.	,000

Varianza total explicada

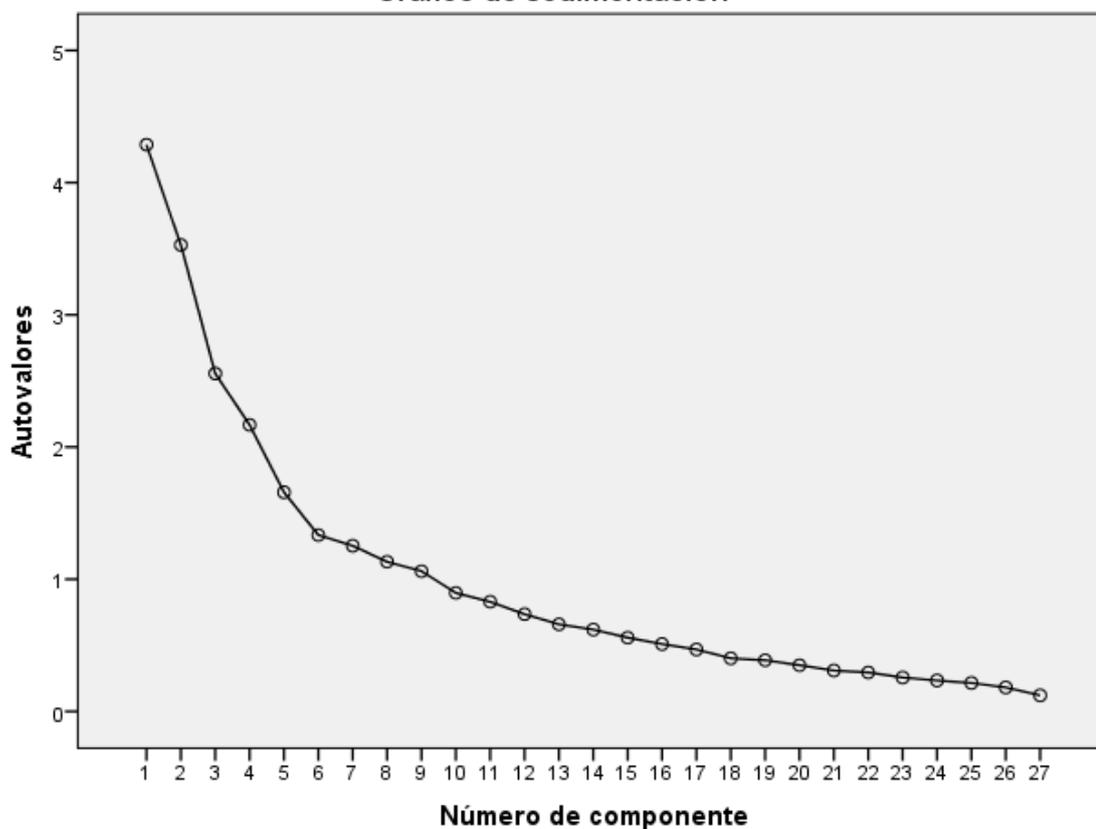
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción	
	Total	% de la varianza	% acumulado	Total	% de la varianza
1	4,287	15,879	15,879	4,287	15,879
2	3,528	13,067	28,946	3,528	13,067
3	2,556	9,468	38,415	2,556	9,468
4	2,168	8,028	46,443	2,168	8,028
5	1,658	6,139	52,582	1,658	6,139
6	1,334	4,940	57,522	1,334	4,940
7	1,253	4,640	62,163	1,253	4,640
8	1,132	4,192	66,354	1,132	4,192
9	1,061	3,929	70,283	1,061	3,929
10	,896	3,320	73,603		
11	,829	3,071	76,675		
12	,735	2,724	79,398		
13	,659	2,439	81,838		
14	,618	2,291	84,128		
15	,558	2,065	86,194		
16	,509	1,886	88,080		
17	,468	1,733	89,813		
18	,401	1,487	91,300		
19	,388	1,436	92,736		
20	,349	1,294	94,029		
21	,310	1,146	95,176		
22	,295	1,092	96,267		
23	,257	,950	97,217		
24	,234	,866	98,084		
25	,215	,797	98,880		
26	,181	,670	99,550		
27	,121	,450	100,000		

ANEXO H (18)

Análisis factorial de los ítems de la encuesta sobre tipos de examen
(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27 y 28)

Varianza total explicada

Componente	Sumas de las saturaciones al cuadrado de la extracción	Suma de las saturaciones al cuadrado de la rotación		
	% acumulado	Total	% de la varianza	% acumulado
1	15,879	3,216	11,910	11,910
2	28,946	2,919	10,809	22,719
3	38,415	2,544	9,424	32,143
4	46,443	2,484	9,200	41,343
5	52,582	1,986	7,354	48,697
6	57,522	1,630	6,037	54,734
7	62,163	1,472	5,452	60,186
8	66,354	1,427	5,285	65,471
9	70,283	1,299	4,812	70,283

Gráfico de sedimentación

ANEXO H (19)

Análisis factorial de los ítems de la encuesta sobre tipos de examen
(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27 y 28)

Matriz de componentes^a

Ítems	Componente								
	1	2	3	4	5	6	7	8	9
5	,627							-,305	
6	,567							-,348	
21	,541	-,486		-,309					
8	,530	-,364		-,406					,327
19	,509					-,356		,450	
9	,503					-,362			,408
15	,466	-,395		-,406					
23	,368	,723							
17		,681	-,370						
20		,626			,423				
10	,337	,611							
3			,621	,394					
24			,590			,314	,335		
22	,531		,581						
16	,456		,523				-,370		
28			-,412	,640					
27	,369	-,349	-,484	,609					
26	,357	-,390	-,444	,533					
2				-,470					
11	,406	,308			-,524	,423			
7	,454				,510				
12	,443	,446			-,487				
13	,326				,476	-,313			
25							,592		,351
4		,350				-,418	,483		
18	,557	-,341						,574	
14					,304			,385	,442

Método de extracción: Análisis de componentes principales.^a

a. 9 componentes extraídos

ANEXO H (19)

Análisis factorial de los ítems de la encuesta sobre tipos de examen
(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27 y 28)

Matriz de componentes rotados^a

Ítems	Componente								
	1	2	3	4	5	6	7	8	9
20	,804								
23	,795								
10	,792								
17	,759								
7	,511							-,351	
8		,807							
21		,785							
15		,702							
2		,457			-,381				
27			,916						
26			,882						
28			,845						
9		,312		,717					
16				,696				-,318	
22				,665	,358				
6				,618					,398
24					,817				
3					,768				
25					,564			,508	
11						,808			
13						-,674			
12	,344			,425		,593			
19				,451			,721		
18		,552					,712		
4								,753	
14									-,736
5		,450							,619

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.^a

a. La rotación ha convergido en 11 iteraciones.