UNIVERSIDAD DE OVIEDO

Tesis Doctoral

Construcción, Desarrollo y Propiedades Psicométricas de una Prueba Práctica de Evaluación de Conductores de Autobús

Autor: D. Pedro Miguel Gallo Álvaro

Departamento: Psicología

Programa: Psicología General

UNIVERSIDAD DE OVIEDO

Tesis Doctoral

Construcción, Desarrollo y Propiedades Psicométricas de una Prueba Práctica de Evaluación de Conductores de Autobús

Autor: D. Pedro Miguel Gallo Álvaro

Director: D. Eduardo García Cueto

Departamento: Psicología

Programa: Psicología Genera

Autorizaciones de lectura de tesis

Resumen

La gran mayoría de las estadísticas actuales muestran que el transporte en autobús es uno de los medios de desplazamiento por carretera mas seguros que existe. Su índice de fallecidos es el más bajo de todos los medios de transporte por carretera. Sin embargo, aún teniendo un índice de riesgo tan bajo, la repercusión social que un único accidente llega a tener es enorme, principalmente debido al elevado número potencial de víctimas que puede suponer.

En las dos últimas décadas, los avances en materia de seguridad en autobuses y autocares, así como la normativa cada vez más restrictiva de acceso a la profesión de conductor de autobús, han permitido un descenso muy destacado en el número de accidentes y de víctimas en este tipo de vehículos.

Aun así, existiendo como existen itinerarios formativos, el acceso a la profesión de conductor de autobús no está aún suficientemente profesionalizado; no existe una formación profesional reglada de acceso a la profesión. La fuerte atomización del sector tampoco ha favorecido que las empresas y empresarios tomen conciencia de la importancia del factor humano en el ámbito de la seguridad en carretera. Actualmente, son las grandes compañías del sector las que, cada vez más, se preocupan por establecer mecanismos que permitan una selección más adecuada y profesionalizada de sus conductores.

Dentro de este contexto, se presenta aquí una prueba práctica de evaluación de conductores de autobús que ha sido elaborada por el departamento de Recursos Humanos de una de las mayores empresas de transporte de viajeros en España y está siendo utilizada como herramienta de selección y formación.

En el presente trabajo, a partir de una muestra de 4.425 pruebas de evaluación, se han estudiado las propiedades psicométricas de dicha prueba práctica. Esta muestra se recopiló a lo largo de un periodo de cuatro años.

Se han estudiado las evidencias de Validez de la prueba basadas en su contenido, así como las evidencias basadas en su estructura interna. También se ha estudiado su relación con otras variables. Se ha estudiado la fiabilidad del instrumento en términos de consistencia interna, calculando el coeficiente α.

El estudio de las propiedades psicométricas de la prueba se completó con un análisis de los ítems tanto desde la perspectiva clásica como desde la Teoría de Respuesta al Ítem, calculando los índices y curvas características correspondientes.

En el presente trabajo se muestran también los resultados de diversos estudios específicos realizados para verificar la relación de cinco variables (*Edad, Experiencia, Sexo, Atención* y *Personalidad*) con los resultados finales de la prueba.

En esta memoria de tesis se incluye también una descripción de la herramienta y un manual de aplicación, elaborado para facilitar el trabajo de los diferentes evaluadores. Asimismo se ha elaborado una tabla general de baremos de los resultados de la prueba.

Las conclusiones del trabajo muestran que el instrumento presenta unas buenas propiedades de medida. Las evidencias de validez basadas en el contenido del test y en su estructura interna son sólidas. Muestra una alta consistencia interna y sus ítems presentan unas características de medida adecuada, tanto en su dificultad como en su capacidad de discriminación.

Se concluye también que, de las cinco variables externas estudiadas en relación con los resultados de la prueba, únicamente la variable *Experiencia* muestra una relación significativa. Los resultados en el test son mejores en los grupos de candidatos que acreditan una mayor experiencia como conductores de vehículos pesados.

Con respecto a las otras variables del estudio no se encontró ninguna relación significativa. Si se realizó un estudio pormenorizado de la variable *Sexo* para comprobar la existencia o no de un funcionamiento diferencial de la prueba en función del género de la persona evaluada. A este respecto se ha podido concluir que no existe un funcionamiento diferencial en ninguno de los trece ítems.

Por último, se presentan y sugieren otras posibilidades de uso de la herramienta, más allá de la selección de personal, y futuras vías de investigación que se abren a raíz de este estudio.

Abstract

The vast majority of current statistics show that bus transportation is one of safest ways to travel by road. Its death rate is the lowest of all road transportation means. However, even with such a low rate, the social impact that a single accident can have is huge, mostly due to the high potential number of victims that it may cause.

In the past two decades, developements in safety on buses and coaches as well as the increasingly restrictive rules of access to the profession of bus driver, have contributed towards a relevant decrease in the number of accidents and victims in such vehicles.

Nevertheless, the access to the bus driver profession is not enough professionalized. There are basic training programmes and on-the-job training, but there are no formal training courses or official qualification, and also the great number of small companies in the market has not favored entrepreneurs to be aware of the relevance of the human factor in road safety. Nowadays, the large companies in the sector are increasingly concerned about this issue, and have started to work thoroughly in developing professional and more accurate drivers selection methods.

Whithin this context, a practical assessment test for bus drivers which has been developed by the Human Resources department of a major passenger transport companies in Spain and which is being used as a tool for selection and training is presented.

In this thesis, based on a sample of 4.425 assessment tests, the psychometric properties of the test have been studied. This sample was collected over a period of four years .

We have studied the evidences of validity based on test content, as well as evidences based on the internal structure of the test. We also studied its relationship with other variables. We have studied the reliability of the instrument in terms of internal consistency by calculating the α coefficient.

The study of the psychometric properties of the test was completed with an item analysis from both the Classical perspective and the Item Response Theory.

In this paper the results of several specific studies to verify the relationship of five variables (Age , Experience, Gender , Care and Personality) with the final test results are also shown .

This PhD dissertation also includes a description of the tool and a use manual, developed to make evaluators' work easier. One general scale of test results, has also been made.

The findings of the study show that the instrument has good measurement properties. The evidences of validity based on test content and internal structure are solid. The test shows high internal consistency and its items have appropiate measure characteristics of both difficulty and discriminating power.

It is also concluded, dealing with the five external variables studied in relation to the test results, only the experience variable shows a significant relationship. The test results are better in groups of more experienced candidates as drivers of heavy vehicles.

For the other variables of the study no significant relationship was found. A detailed study of the variable Gender was performed to check the existence of a differential item functioning by gender of the person being assessed.

Finally, we present and suggest other ways of using the tool beyond recruitment, and future research avenues that are open as a result of this study.

Aviso de propiedad intelectual

La prueba de evaluación de conductores aquí descrita ha sido utilizada con autorización expresa de la sociedad Automóviles Luarca S.A. propietaria de los derechos de su propiedad intelectual, para fines exclusivamente de investigación y no deberá ser utilizada o reproducida, en su totalidad o en parte, por ningún otro particular o empresa sin autorización de su propietaria.

Índice de Contenidos

Índice de contenidos xvii

Índice de tablas xxi

Índice de figuras xxiii

Capítulo 1: **El problema de la Seguridad Vial y la implicación del sector de Transporte de Viajeros por Carretera** **1**

1.1 Perspectiva histórica de la accidentalidad en carretera 2

1.2 Europa: Una década para salvar 100.000 vidas 8

1.3 Medidas de Seguridad Vial en España 15

1.4 La siniestralidad en autobús y el sector del transporte de viajeros por carretera 22

Capítulo 2: **Factor Humano y Selección de Conductores** **35**

2.1 Antecedentes históricos: Teorías y modelos explicativos 36

2.1.1 Modelos de habilidades 39

2.1.2 Modelos comportamentales 42

2.2 El modelo español de Selección de conductores 48

2.3 ¿Por qué es necesaria una evaluación práctica de los conductores profesionales? 55

Capítulo 3: **Construcción, Descripción y Condiciones Generales para la aplicación de la Prueba** **61**

3.1 Construcción y descripción de la Prueba Práctica de Evaluación de Conductores 62

3.1.1 Ficha Técnica 62

3.1.2 Criterios de elaboración 63

3.1.3 Significación de cada ítem 66

3.1.3.1 Control de las revoluciones 67

3.1.3.2 Uso de retarder 67

3.1.3.3 Uso del freno de servicio 68

3.1.3.4 Anticipación 68

3.1.3.5 Control emocional 69

3.1.3.6 Observación de los retrovisores 69

3.1.3.7 Distancia de seguridad 70

3.1.3.8 Manejo del volante 71

3.1.3.9 Atención a la señalización vial 71

3.1.3.10 Fluidez en la circulación 72

3.1.3.11 Observación del cuadro de mandos 72

3.1.3.12 Manejo del cambio 73

3.1.3.13 Sentido de la maniobrabilidad 73

3.2 Condiciones generales de aplicación 75
3.2.1 Indicaciones sobre el desarrollo de la prueba 75
3.2.2 Normas de aplicación 77
3.2.3 Puntuación 82
3.3 Descripción de otras pruebas utilizadas en el estudio 83
3.3.1 Para el estudio sobre atención 83
3.3.2 Para el estudio sobre personalidad 84
3.4 Planteamiento de hipótesis 87
3.4.1 Estado de la cuestión 87
3.4.2 Objetivos del estudio 91
3.4.3 Hipótesis de trabajo 92

Capítulo 4: **Propiedades Psicométricas de la Prueba: Fiabilidad y Validez** **95**
4.1 Descripción general de la muestra 96
4.2 Análisis de Validez 103
4.2.1 Evidencias basadas en el contenido del test 105
4.2.2 Evidencias basadas en la estructura interna del test 106
4.2.3 Evidencias basadas en las relaciones con otras variables 115
4.2.4 Resultados 117
4.2.4.1 Evidencias de validez de contenido 117
4.2.4.2 Evidencias de validez factorial 119
4.2.4.3 Otras evidencias de validez 128
4.3 Análisis de Fiabilidad 131
4.3.1 Modelo Lineal Clásico 131
4.3.2 Coeficiente Alpha (α) 132
4.3.3 Resultados desde la Teoría Clásica de los Test 133
4.4 Análisis de los ítems 135
4.4.1 Fundamentos del Análisis de ítems 135
4.4.2 Perspectiva clásica del Análisis de ítems 136
4.4.3 Resultados 139
4.5 Teoría de Respuesta a los Ítems 143
4.5.1 Principios generales 143
4.5.2 Análisis de ítems desde la TRI 146
4.5.3 Resultados desde la TRI 147
4.6 Cálculo de la tabla de baremos 152

Capítulo 5: **Estudios Específicos** **157**
5.1 Edad y resultados en la prueba 158
5.1.1 Edad y conducción 158
5.1.2 Estudio y resultados sobre Edad 160

5.2 Efecto de la experiencia sobre los resultados en la prueba 169
5.2.1 Experiencia y conducción 169
5.2.2 Estudio y resultados sobre Experiencia 172
5.3 Sexo y resultados en la prueba: Estudio de sesgo 179
5.3.1 Sexo y conducción 179
5.3.2 Funcionamiento Diferencial de los Ítems 182
5.3.3 Estudio y resultados sobre DIF 184
5.4 Atención y resultados en la prueba 188
5.4.1 Atención y conducción 188
5.4.2 Estudio y resultados sobre Atención 190
5.5 Personalidad y resultados en la prueba 198
5.5.1 Personalidad y conducción 198
5.5.2 Estudio y resultados sobre Personalidad 201

Capítulo 6: **Discusión de los Resultados** **205**
6.1 Discusión de las propiedades Psicométricas 206
6.1.1 Evidencias de validez 207
6.1.2 Fiabilidad 213
6.1.3 Análisis de ítems 214
6.2 Discusión de los estudios específicos 217
6.2.1 Edad 217
6.2.2 Experiencia 218
6.2.3 Sexo 219
6.2.4 Atención 220
6.2.5 Personalidad 221

Capítulo 7: **Conclusiones y potencialidades futuras de la Herramienta** **223**
7.1 Conclusiones 224
7.2 Posibilidades de uso 226
7.3 Futuras vías y proyectos de investigación 228

Referencias bibliográficas **231**

Apéndice: **Anexos, Tablas y Análisis Estadísticos** **243**
8.1 Salida completa de indicadores de ajuste del Análisis Factorial Confirmatorio 244
8.2 Resultados Multilog y curvas características de los ítems 246
8.3 Contraste de Levene calculado en el estudio sobre Experiencia 256
8.4 Resultados de las escalas del d2 en el estudio de Atención 257
8.5 Tabla de distribución de frecuencias de puntuaciones del NEO 263
8.6 Análisis de varianza completo del estudio de personalidad 266

Índice de Tablas

Capítulo 1: **El problema de la Seguridad Vial y la implicación del sector de Transporte de Viajeros por Carretera**

Tabla 1 Fallecidos y accidentes con víctimas en España 1970-2009 6

Tabla 2 Evolución de cifras de accidentalidad en España 2001-2010 16

Tabla 3 Accidentalidad de autobuses en España por área en que se producen (2010) 25

Tabla 4 Condiciones para la obtención de los permisos de autobús. R.D. 818/2009 29

Capítulo 4: **Propiedades Psicométricas de la prueba: Fiabilidad y Validez**

Tabla 5 Estadísticos descriptivos de la puntuación del test para la muestra general 98

Tabla 6 Distribución de la muestra según la edad 100

Tabla 7 Distribución de la muestra según la antigüedad del permiso 102

Tabla 8 Condiciones necesarias para realizar un AFC 113

Tabla 9 Varianza explicada basada en autovalores, submuestra 1 (N=1929) 121

Tabla 10 Matriz factorial no rotada, submuestra 1 (N=1929) 123

Tabla 11 Matriz factorial rotada, submuestra 1 (N=1929) 123

Tabla 12 Varianza total explicada. AF 2º orden, submuestra 1 (N=1929) 124

Tabla 13 Matriz de correlación entre factores, submuestra 1 (N=1929) 125

Tabla 14 Indicadores de ajuste del modelo de AFC, submuestra 2 (N=1929) 127

Tabla 15 Tabla de clasificación observada (N=390) 130

Tabla 16 Tabla de clasificación pronosticada (N=390) 130

Tabla 17 Resultados de α en FACTOR, escala completa 134

Tabla 18 Valor de α en las tres escalas de la prueba de evaluación 134

Tabla 19 Correlación recomendada en función del nivel de medida 138

Tabla 20 Estadísticos descriptivos de los ítems 140

Tabla 21 Estadísticos total-elemento 141

Tabla 22 Índices de fiabilidad de los elementos 142

Tabla 23 Parámetros TRI de los ítems 148

Tabla 24 Baremo de puntuaciones. Muestra general 154

Capítulo 5: **Estudios Específicos**

Tabla 25 Grupos de Edad 161

Tabla 26 Análisis de varianza para Puntuación en el test por Edad 162

Tabla 27 Comparaciones Post Hoc. Scheffe para Edad 164

Tabla 28 Matriz de tamaño del efecto (*g* de Hedges para EDAD-PUNTUACIÓN) 168

Tabla 29 Distribución de los grupos por experiencia profesional con vehículo pesado 173

Tabla 30 Análisis de varianza para Puntuación en el test por Experiencia 173

Tabla 31 Comparaciones Post Hoc. C de Dunnett para Experiencia 174

Tabla 32 Matriz de Tamaño del efecto (*g* de Hedges para EXPERIENCIA-PUNTUACIÓN) 175

Tabla 33 Análisis de varianza para Puntuación en el test por Edad con Experiencia como covariable 176

Tabla 34 Comparaciones por pares. Ajuste Bonferroni para Edad con Experiencia como covariable 177

Tabla 35 Prueba T de comparación de medias para la variable sexo 185

Tabla 36 Resultados del estudio de DIF 187

Tabla 37 Estadísticos descriptivos Total Conducción (N=106) 191

Tabla 38 Estadísticos descriptivos Prueba de Atención d2 (N=106) 191

Tabla 39 Análisis de Varianza para puntuación en el test por nivel atencional TR 192

Tabla 40 Resumen de ANOVA para escalas d2 193

Tabla 41 Tabla de contingencia Puntuación total en test por total respuestas en test de atención 194

Tabla 42 χ^2 de Pearson Puntuación total en test por total respuestas en el test de atención 195

Tabla 43 Coeficiente Eta total test por total respuestas 196

Tabla 44 Resumen de valores χ^2 de Pearson y Eta para tablas de contingencia 197

Tabla 45 Estadísticos descriptivos NEO PI-R 202

Tabla 46 Distribución de puntuaciones agrupadas, Neuroticismo NEO PI-R 202

Tabla 47 Análisis de Varianza para puntuación en el test por resultados en test NEO PI-R de personalidad 203

Índice de Figuras

Capítulo 1: **El problema de la Seguridad Vial y la implicación del sector de Transporte de Viajeros por Carretera**

Figura 1 Fallecidos en carretera por millón de habitantes (2001) 9
Figura 2 Porcentaje de reducción en el número de fallecidos en carretera entre 2001 y 2010 12
Figura 3 Fallecidos en carretera por millón de habitantes (2009) 13
Figura 4 Evolución del número de fallecidos en accidente de tráfico 1980-2010 17
Figura 5 Porcentaje de uso de cinturón de seguridad en turismos y furgonetas CARRETERA/ZONA URBANA 19
Figura 6 Controles de velocidad realizados por la Agrupación de Tráfico de la Guardia Civil 20
Figura 7 Controles de alcoholemia realizados por la Agrupación de Tráfico de la Guardia Civil 21
Figura 8 Distribución de fallecidos por tipo de vehículo (UE 2009) 24
Figura 9 Anverso y reverso de tarjeta CAP 31
Figura 10 Hoja de registro analógico y tarjeta de tacógrafo digital 32

Capítulo 2: **Factor Humano y la Selección de conductores**

Figura 11 Fuentes de factores de riesgo en la conducción 37
Figura 12 Porcentaje de implicación de las diferentes causas de accidentes de tráfico 37

Capítulo 3: **Construcción, Descripción y Condiciones generales para la aplicación de la prueba**

Figura 13 Ejemplo de indicación de dirección durante la prueba 80

Capítulo 4: **Propiedades Psicométricas de la prueba: Fiabilidad y Validez**

Figura 14 Histograma de distribución de la puntuación total del test 99
Figura 15 Distribución del número de participantes por provincia 101
Figura 16 Ejemplo de Gráfico de Sedimentación 109
Figura 17 Ejemplo de Path Diagram de AFC 112
Figura 18 Matriz de correlaciones policóricas, submuestra 1 (N=1929) 121
Figura 19 Adecuación de la matriz de correlaciones, submuestra 1 (N=1929) 121
Figura 20 Análisis paralelo, submuestra 1 (N=1929) 122
Figura 21 Gráfico de sedimentación. AF 2º orden, submuestra 1 (N=1929) 124
Figura 22 Path Diagram para AFC, submuestra 2 (N=1928) 126
Figura 23 Función de información de la prueba 147
Figura 24 CCI y Función de Información ítem 11 149
Figura 25 CCI y Función de Información ítem 7 150
Figura 26 CCI y Función de Información ítem 9 150

Figura 27 CCI y Función de Información ítem 13 151

Capítulo 5: **Estudios Específicos**

Figura 28 Gráfico de medias marginales estimadas. ANOVA Edad 165

Figura 29 Gráfico de medias marginales estimadas. ANOVA Experiencia 175

Figura 30 Gráfico de medias marginales estimadas. ANOVA Edad con Experiencia como covariable 178

Figura 31 Representación de un ítem con DIF Uniforme 183

Figura 32 Representación de un ítem con DIF No Uniforme 184

Figura 33 Distribución de hombres y mujeres 185

Figura 34 Gráfico de medias marginales estimadas ANOVA nivel TR 193

Figura 35 Gráficos de medias marginales estimadas, Neuroticismo (N) y Responsabilidad (C) 204

En este mundo hay cosas que solo puedes hacer por ti mismo y cosas que solo puedes hacer con otra persona. Es importante ir combinando las unas con las otras.

Haruki Murakami (1949-)

1 El problema de la Seguridad Vial y la implicación del sector de Transporte de Viajeros por Carretera

En este capítulo se muestran algunos datos que contextualizan la cuestión de la Seguridad Vial, tanto en España como en el mundo, tratándola como un problema de Salud Pública. Se describen estadísticas, acciones y medidas aplicadas, no solo en nuestro país sino también en otros, principalmente europeos, que ofrecen una visión inicial de conjunto.

Esta introducción da pie a la problemática del autobús, como medio de transporte, integrándola en un contexto global. Se muestran estadísticas específicas que señalan los puntos más característicos del entorno empresarial del Transporte de Viajeros por Carretera en España; así como normativas específicas en relación con la Seguridad Vial y los estudios más relevantes.

1.1 Perspectiva Histórica de la Accidentalidad en Carretera

Probablemente, el siglo XX ha visto sido escenario de los mayores avances tecnológicos desde que el hombre habita la Tierra. Si no es así, lo que es seguro es que estos avances han sido los más rápidos. Tuvieron que pasar doscientos mil años desde la aparición del *homo sapiens* hasta que, en diciembre de 1903, los hermanos Wrigth hicieran posible el primer vuelo a motor con una máquina fabricada por el hombre, recorriendo apenas cuarenta metros en doce segundos . Solo sesenta y seis años después, Neil Amstrong daba un *gran paso para la humanidad* dejando su huella sobre el suelo de la luna, tras un viaje de 384.400 kilómetros en seis días.

Desde la invención de la rueda, allá por el 4.000 A.C., hubo que esperar hasta el año 1908 para que el uso del automóvil se generalizara con la salida al mercado del famoso Ford Modelo T. Tras 105 años, el automóvil y el sistema de tráfico han crecido y evolucionado tanto que ya forman parte del modo en que el ser humano se relaciona con su entorno.

Ambos desarrollos aquí comentados tienen un denominador común: para su eclosión y progreso definitivo requerían, de una fuente de energía adecuada, que proporcionara potencia suficiente sin ocupar demasiado espacio. El motor de combustión interna, en 1885, acabó por imponerse a sus rivales, tanto por eficiencia como por tamaño, y trajo consigo el cambio definitivo de paradigma (Eckermann, 2001).

Esta nueva forma de desplazarse y relacionarse, a nivel socioeconómico ,tuvo una influencia directa en la gran mayoría de los avances y progresos de la humanidad en el siglo pasado y, aún hoy, continúa teniéndola en nuestro presente y

en nuestro horizonte más inmediato; pero no está exenta de problemas y consecuencias perjudiciales para nuestro futuro. No en vano, uno de esos problemas ya se manifestaba en los albores de esta tecnología.

En el año 1769, Nicolas Joseph Cugnot (1725-1804), un ingeniero militar del ejército francés, construyó, por indicación del Ministro de Exteriores, la primera versión de un transporte de tres ruedas, guiado por un timón delantero y con capacidad para cuatro personas (Ziel, 2003). Aquel vehículo, concebido originalmente como un aparato destinado al transporte y arrastre de mercancías, presentaba una particularidad que lo distinguía del resto y que haría que la Historia lo reconociese como un pionero. Se trataba de un vehículo autopropulsado por un motor de vapor, de modo que se convirtió en el primer vehículo automóvil de la Historia.

Unos años antes, Cugnot había conseguido diseñar un mecanismo, por medio de un sistema de trinquete y cadena, capaz de convertir el movimiento recíproco de unos pistones impulsados por vapor en un movimiento rotatorio, permitiendo incluso el desplazamiento marcha atrás. La incorporación de este mecanismo a un vehículo de ruedas llevó a crear un transporte que el ingeniero llamó *Fardier à vapeur* (carro de vapor). En 1770 fabricó una versión mejorada del prototipo original, con el objetivo de que se usara como vehículo militar de carga y arrastre. Este furgón de artillería medía 7,25 metros de largo y 2,19 de ancho; estaba dotado con una caldera en la parte delantera de 1,5 metros de diámetro y tenía, según su creador, capacidad para transportar hasta cuatro toneladas, pero este dato nunca pudo ser comprobado (Eckermann, 2001). Al año siguiente, mientras se realizaban unas pruebas y se enseñaba el manejo de la máquina a otros oficiales del ejército francés, uno de sus conductores perdió el control y estrelló el vehículo contra un muro de ladrillo. El primer automóvil de la historia acababa de sufrir el primer accidente de tráfico del que se tiene constancia (Ziel, 2003).

Lo que vino a continuación cumplió este funesto augurio. La generalización del uso del motor de combustión interna fue la chispa que alumbró el crecimiento exponencial del uso del automóvil y, por extensión, de los accidentes de tráfico.

Un siglo después de la aventura de Cugnot, en el año 1886, solamente dos meses más tarde de la introducción oficial del automóvil en Estados Unidos, se produjo el primer accidente del que se tiene registro en ese país (Ziel, 2003). Se trató de una colisión entre un automóvil y una bicicleta. El ciclista se recuperó sin

problemas y el conductor pasó esa noche en el calabozo, pero, ya en aquel momento, se sentaron las bases de la difícil relación actual entre ambos medios de transporte.

Tres años después de este incidente se registró el primer accidente con víctima mortal en el país norteamericano. No se trató del primero a nivel mundial, ya que existen registros de accidentes con víctimas en otros países en fechas anteriores, sobre todo en Gran Bretaña. El 13 de septiembre de 1899, en la ciudad de Nueva York, el conductor de un coche atropelló a un peatón que cruzaba en ese momento por la calle, causándole la muerte. Diez años más tarde, en Estados Unidos ya morían 1.000 personas anualmente a consecuencia de algún tipo de accidente de tráfico. Solamente nueve años después, en 1918, se alcanzó la cifra global de 10.000 fallecidos en carretera. Para completar el panorama de los primeros treinta años del siglo, en el año 1930 ya habían fallecido 30.000 personas en accidentes de este tipo. La situación generada por el crecimiento tan desmesurado del número de accidentes y fallecidos hizo que las autoridades comenzaran a preocuparse por este tema. El resultado fue la introducción de diversas medidas, pioneras en los aspectos relacionados con la seguridad, como los frenos hidráulicos, que venían a sustituir a sistemas de frenado muy rudimentarios o incluso inexistentes, o los cristales de vidrio laminado (Ziel, 2003).

A pesar de todas estas acciones no se lograría disminuir la siniestralidad en carretera ni minimizar los efectos de los accidentes en términos de fallecidos o heridos graves. La mejora y desarrollo de los vehículos, en aspectos como la potencia o el rendimiento, propiciaron un aumento de la velocidad. Sin embargo no existió un desarrollo parejo de las medidas de seguridad en los vehículos y carreteras; esto trajo consigo la imposibilidad de controlar el creciente número de accidentes en los años venideros (Ziel, 2003).

La década de los cincuenta es un periodo clave en Estados Unidos. En 1951, se alcanza la terrible cifra de un millón de muertos en accidentes de tráfico (Norman, 1963). Con una media anual de 35.000 muertes en carretera, se comienza a asumir, definitivamente, que el problema de la accidentalidad del tráfico es una amenaza real para la Salud Pública. Se estaba diezmando silenciosamente a toda una generación, restándole años de productividad al país y generando una catástrofe social y económica.

En España, el desarrollo de los acontecimientos no fue muy diferente. A principios de siglo, no era extraño leer reseñas en las secciones de sucesos de los periódicos que informaran de percances o colisiones de tráfico. En su mayoría se trataba de atropellos o salidas de vía. El 26 de mayo de 1908 el periódico ABC publicó, en la página 10, la noticia del que bien pudiera ser el primer accidente de circulación con víctima mortal en España. El periodista, del servicio particular del periódico en Sevilla, narra la noticia describiendo la *Catástrofe automovilística* de la siguiente manera:

> [...] La (versión) que parece más cercana a la verdad dice que los expedicionarios llevaban una velocidad extraordinaria, y al llegar a una vuelta de la carretera, en el sitio llamado la Moncloa, entre los kilómetros 480 y 481, no pudo el que guiaba moderar el ímpetu del carruaje para virar convenientemente, cayendo el automóvil por un terraplén de no mucha altura, pero de forma en que el vehículo volcó, cogiendo debajo a los viajeros. [...] (Anónimo, 1908a pag. 10)

El fallecido resultó ser el Marqués de las Cuevas del Becerro, del que el propio periódico hace una loa elevada, al estilo de la época. Casualmente, unas pocas páginas antes, la misma edición del periódico se hace eco de un atropello bajo el titular *Los automóviles*. El reportero inicia su narración con la frase "El atropello del día..." (Anónimo, 1908b pag. 8). Esto da idea de lo frecuentes que resultaban, ya en aquella época, los incidentes provocados por problemas de adaptación de los peatones a la presencia de estos nuevos vehículos con los que habían empezado a convivir. El joven encuadernador Ricardo Bueno Ruiz, que así se llamaba el atropellado, tuvo la suerte de salir de aquel incidente únicamente con una fractura de pierna y algunas abrasiones en la cara; el buen Marqués de las Cuevas, desgraciadamente, no pudo decir lo mismo.

El control administrativo de los vehículos de motor comenzó en España en el año 1900; en 1907 se establecen las normas generales para la matriculación de vehículos. Aún así, durante las tres primeras décadas del siglo XX, los vehículos de motor eran artículos de lujo solo accesibles a familias adineradas, con capacidad para la compra y mantenimiento de este tipo de transportes. El paso por la Guerra Civil (1936-1939) y el periodo de postguerra tampoco favorecieron el desarrollo y generalización de este medio de transporte (Rodríguez, 1987).

Habría que esperar hasta la década de los años cincuenta para observar un verdadero desarrollo y expansión del automóvil, como medio de transporte. En ese

periodo se crean, SEAT, que posteriormente fabricó con licencia FIAT; FASA, que fabricaba vehículos RENAULT y Barrerios Diesel, industria especializada en la construcción de autocamiones. Todas continuaban el camino marcado en 1946 por ENASA y sus *Pegaso*, empresa pionera (Rodríguez, 1987).

Este periodo de crecimiento concluye con la Ley 80 de 1959, en virtud de la cual se crea la Jefatura Central de Tráfico, como órgano de ordenación y coordinación de las funciones de vigilancia y regulación de tráfico, y también la Guardia Civil de Tráfico (Rodríguez, 1987). Por aquel entonces se matriculaba en España un promedio anual de 258.000 vehículos y se comenzaron a registrar estadísticas de accidentalidad más completas.

Total	1970	1980	1990	2000	2005	2008	2011
Fallecidos	5.456	6.522	9.032	5.776	4.442	3.100	2.060
Accidentes con víctimas	57.968	67.803	101.507	101.729	91.187	93.161	83.027

Tabla. 1: Fallecidos y accidentes con víctimas en España 1970-2011
Fuente IRTAD, DGT (elaboración propia)

Ya en el año 1970, el número de fallecidos en accidentes en carretera se situaba en torno a los 5.500 (Tabla 1), en casi 60.000 accidentes con víctimas.

Entre los años 1983 y 1989 el parque móvil de vehículos en España se incrementó en un 33%. De la misma manera, el volumen de tráfico en carreteras aumentó un 41%. En este contexto, durante este periodo, el número de fallecidos en accidente de tráfico creció de manera continuada hasta alcanzar, en 1989, un pico de 9.344 personas fallecidas (IRTAD, 2010).

A partir de ese punto, donde la cifra se acercaba a valores realmente intolerables, la introducción de nuevas medidas y de una reforma de la Ley de Tráfico favoreció el paulatino descenso del número de muertes en carretera. En el periodo de los cuatro años siguientes, hasta 1994, el dato descendió un 39%, con respecto a la cifra de 1989. El volumen de tráfico continuaba creciendo, pero se realizó una gran inversión en la construcción y mejora de las vías de gran capacidad (autopistas y autovías).

Este ritmo de descenso positivo se truncó en el siguiente periodo. Entre 1995 y 2003 no se puede identificar una tendencia clara, habiendo muchos altibajos en el

número de fallecidos. El volumen de tráfico interurbano creció de manera significativa, un 76%. Sin embargo, en el año 2003 había construidos y en funcionamiento 4.000 kilómetros más de vías de alta capacidad que en 1994, 10.296 km en 2003, frente a los 6.497 km de mediados de los noventa.

A partir del año 2004, el descenso de fallecidos es muy significativo. El número de muertes se reduce, hasta 2010, en un 55% (IRTAD, 2010). En la actualidad, el último dato aportado por IRTAD indica que el descenso total, en el periodo 2000-2011, es de un 64% (IRTAD, 2013). Siguiendo la guía marcada desde los principales organismos europeos, orientada a la reducción de la accidentalidad en los países de la Unión, se pone en marcha en España el Plan Estratégico de Seguridad Vial.

Tras este recorrido de algo más de cien años, en la última década Europa ha venido marcando el paso y España ha respondido con solvencia.

1.2 Europa: Una década para salvar 100.000 vidas

El Consejo Europeo de Seguridad en el Transporte (European Transport Safety Council [ETSC]) es una organización no gubernamental, de ámbito internacional, creada en 1993 con el objetivo de servir como fuente imparcial de consejo en materia de seguridad vial a los principales organismos administrativos europeos, así como a sus gobiernos (ETSC, 2009). Nace como respuesta al creciente número de víctimas que se venía produciendo anualmente en las carreteras europeas y a una situación que se estaba convirtiendo en insostenible. En el año 1993, cuando se crea el Consejo, la cifra de fallecidos en Europa, a consecuencia de accidentes de tráfico, era superior a los 65.000[1]. Para cualquier sociedad, que aspira a ser avanzada, civilizada y segura, éste es un peaje intolerable.

Siete años después, a finales de 2000, la situación no era mucho mejor; más de 56.000 muertos en carretera[2] no eran, ni mucho menos, el indicador de que se estuviera yendo en la dirección correcta para atajar esta lacra (ETSC, 2000). La figura 1 ilustra la situación de los diferentes países europeos en el año 2001. La mayoría de ellos estaban en una cifra anual superior a los cien fallecidos por millón de habitantes, la media europea se situaba en 112 fallecidos por millón e, incluso, en países como Lituania y Letonia el dato traspasa la barrera de los 200 muertos por

[1] 65.441 Fallecidos. (Fuente CARE EU road accidents database)

[2] 56.427 Fallecidos. (Fuente CARE EU road accidents database)

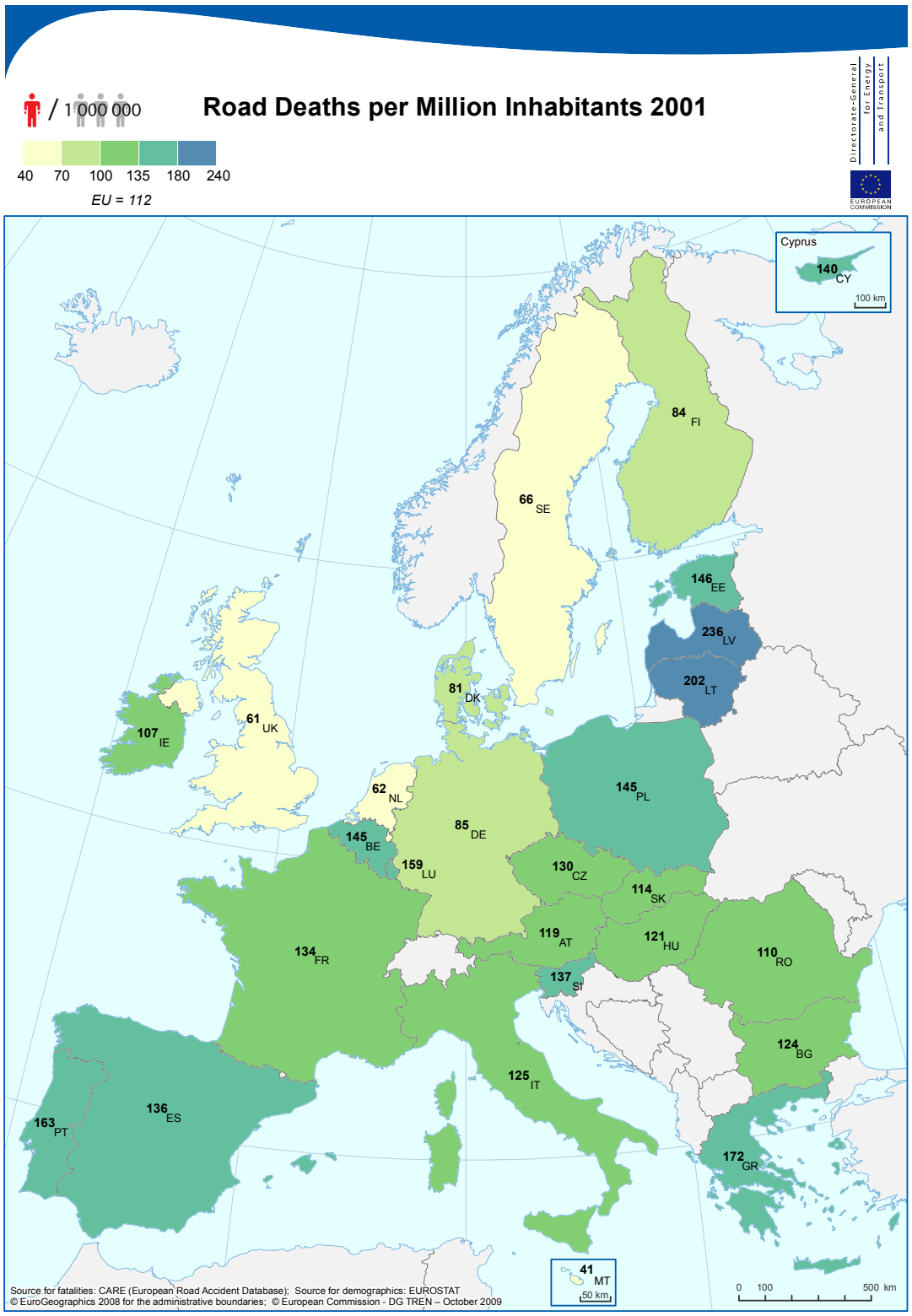


Fig. 1: Fallecidos en carretera por millón de habitantes (2001)
Tomado de CARE (EU road accidents database)

millón. Únicamente los Países Bajos, Suecia y Reino Unido se mueven en valores inferiores a los setenta fallecidos por millón de habitantes. En España, la tasa de mortalidad es de 5.517 muertos en accidentes de tráfico, lo que arroja la cifra de 136 fallecidos por millón de habitantes. Eso implica que, durante el año 2000, cada día morían en España 15 personas como resultado de un accidente de circulación. En

palabras de la, en aquel momento, Vicepresidenta de la Comisión de Transporte de la Unión Europea, Dña. Loyola de Palacio:

> *Simplemente no podemos resignarnos a una tragedia diaria de semejante magnitud. De forma individual y de forma colectiva, cada parte de la Sociedad debe desempeñar su papel para promover la Seguridad en Carretera*. (ETSC, 2000)[3]

Ante este panorama, la Unión Europea decidió establecer, para el periodo 2001-2010, un objetivo cuantificable y directamente relacionado con la seguridad en carretera, siguiendo un planteamiento similar al que ya existía en el Reino Unido. En el año 2001 se pone encima de la mesa el objetivo, para toda la Unión Europea, de reducir en un 50% la mortalidad en carretera para un periodo de diez años (ETSC, 2010b). A grandes rasgos, esto suponía pasar de 54.000 a 27.000 fallecidos anuales; un planteamiento, sin duda, ambicioso y que iba a requerir de la máxima implicación y colaboración por parte de todos los gobiernos de la Unión.

Por primera vez las acciones de la Unión Europea encaminadas a mejorar la seguridad en carretera se dirigían a un objetivo concreto, cuantificable e, inesperadamente, mucho más complejo y retador de lo que cabía esperar, (ETSC, 2010b).

En primer lugar, la entrada a la Unión, en el año 2004, de diez nuevos países, con índices de siniestralidad y mortalidad sensiblemente superiores a los de la Europa de los 15, no estaba prevista cuando se fijó el objetivo de reducción.

En segundo lugar, se hizo patente que la mera fijación del objetivo no podía, de ninguna manera, asegurar la consecución del mismo. La Unión Europea debía tomar, por si misma, las riendas del proceso, proporcionando una guía de acciones y medidas que los países miembros pudieran poner en práctica para reducir, de esta forma, las muertes en accidentes de tráfico más rápido de lo que se venía haciendo hasta ese momento.

Por último, estas acciones carecerían de valor si no se veían acompañadas de una legislación fuerte que las sustentara y permitiera actuar, de manera diligente e inapelable, sobre aquellos riesgos de la carretera que producían mayor número de muertes.

---

[3] *We simply could not resign ourselves to a daily tragedy on the current scale. Individually and collectively, all sections of society must play their part in promoting road safety.* En el original (N. del A.)

El proceso se inició de manera lenta, a mitad de camino, en el año 2005, los países que encabezaban la reducción de las muertes en carretera, principalmente **Francia**, **Luxemburgo** y **Bélgica**, rondaban el 30% de descenso en el número de fallecidos. La mayoría no llegaba al 20% e incluso algunos países como **Hungría**, **República Checa** o **Lituania**, no solo no habían logrado reducir su número de fallecidos en carretera sino que este dato había aumentado en el último año. **España** aún estaba situada por encima de los cien fallecidos por millón de habitantes.

Así las cosas, el año 2006 marca un cambio en las políticas de los países y en la visión de este problema. La aplicación de nuevas medidas de Seguridad Vial y la introducción de nuevos instrumentos y políticas pasa a primer plano. Aquellos países que llevaban varios años trabajando en la introducción de este tipo de Planes y Medidas de Seguridad Vial comenzaron, por fin, a notar los efectos positivos en su camino hacia el objetivo propuesto. El número de fallecidos estaba descendiendo, pero era posible progresar más rápido y a mejor ritmo (Elvik y Vaa, 2006).

La información del final del proceso está disponible en el quinto informe PIN (Road Safety Performance Index) (ETSC, 2011). En él se presentan resumidos los resultados de todo el proceso de implantación de medidas de Seguridad Vial durante estos últimos diez años (Fig. 2).

Es indudable, a la luz de los datos, que se ha hecho un gran avance en todo el territorio de la Unión Europea, en materia de seguridad. **Letonia**, **Estonia**, **Lituania**, **España**, **Luxemburgo**, **Suecia**, **Francia** y **Eslovenia** son los ocho países que han conseguido alcanzar el objetivo de reducción marcado, señalados en verde oscuro en la figura 2. Los países representados en verde claro, encabezados por **Portugal**, con un porcentaje muy próximo a conseguir el objetivo del 50%, han tenido un descenso igual o mayor que la media europea, que se sitúa en un 43% de reducción total para los 27 países de la Unión, siendo de un significativo 47% la media de los 15 países que iniciaron el camino en el año 2001. Otros, representados en naranja y rojo, han progresado en el descenso en un porcentaje inferior, pero, en mayor o menor medida todos han logrado una reducción significativa en los niveles de mortalidad .

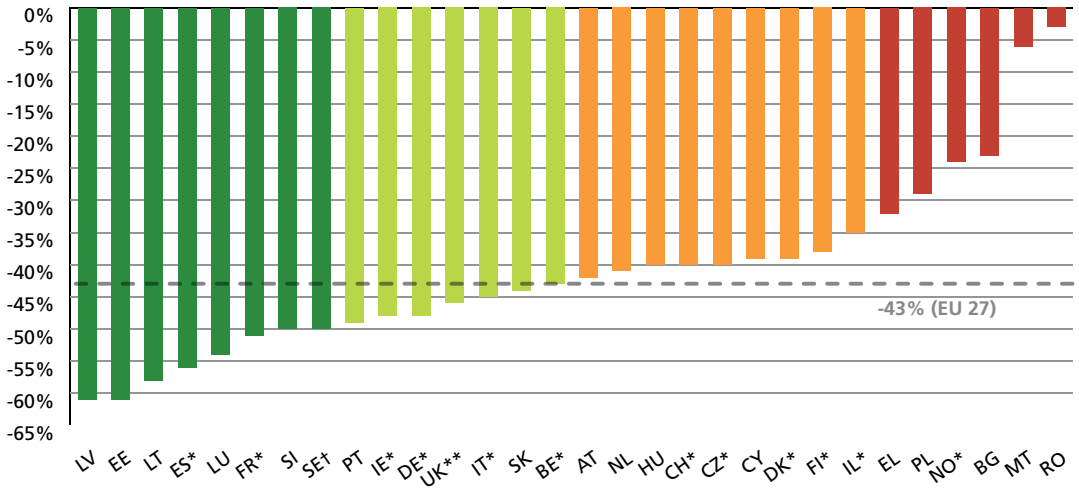


Fig. 2: Porcentaje de reducción en el número de fallecidos en carretera entre 2001 y 2010
* Estimaciones provisionales al carecer de las cifras definitivas del año 2010 cuando se imprimió el informe.
** Las estimaciones para el Reino Unido han utilizado el Indicador Rápido Europeo de CARE.
† Para Suecia la definición de muerte en carretera ha cambiado y los suicidios no son tenidos en cuenta. Se ha ajustado la serie temporal de tal forma que los datos de años precedentes los excluyan también.
Tomado de ETSC (5th PIN Report)

El establecimiento de un objetivo concreto, medible y verificable se ha mostrado como una medida apropiada para impulsar la mejora y desarrollo de planes de Seguridad Vial en los países[4].

Como se puede observar en la figura siguiente (Fig. 3), en una década, el tono del mapa europeo se ha vuelto más claro. Ya no predominan los países por encima de los cien fallecidos por millón de habitantes. La media europea ha descendido a 70 fallecidos por millón. Los países que en 2001 superaban los 100 fallecidos por millón han reducido sensiblemente esa medida y aquellos que superaban los 200 fallecidos por millón ya están situados en valores inferiores, cercanos a los de sus países vecinos, lo que acredita un descenso muy significativo en el número anual de muertes.

En el caso español, los datos correspondientes a 2010 arrojan un dato aún más positivo, en el periodo de 2009 a 2010, se ha pasado de los 59 fallecidos por millón de habitantes, que figuran en el mapa aquí representado, a 54.

[4] Abreviaturas de los países: (LV Letonia; EE Estonia; LT Lituania; ES España; LU Luxemburgo; FR Francia; SL Eslovenia; SE Suecia; PT Portugal; IE Rep. Irlanda; DE Alemania; UK Reino Unido; IT Italia; SK Eslovaquia; BE Belgica; AT Austria; NL Países Bajos; HU Hungría; CH Suiza; CZ Rep. Checa; CY Chipre; DK Dinamarca; FI Finlandia; IL Luxemburgo; EL Grecia; PL Polonia; NO Noruega; BG Bulgaria; MT Malta; RO Rumanía).

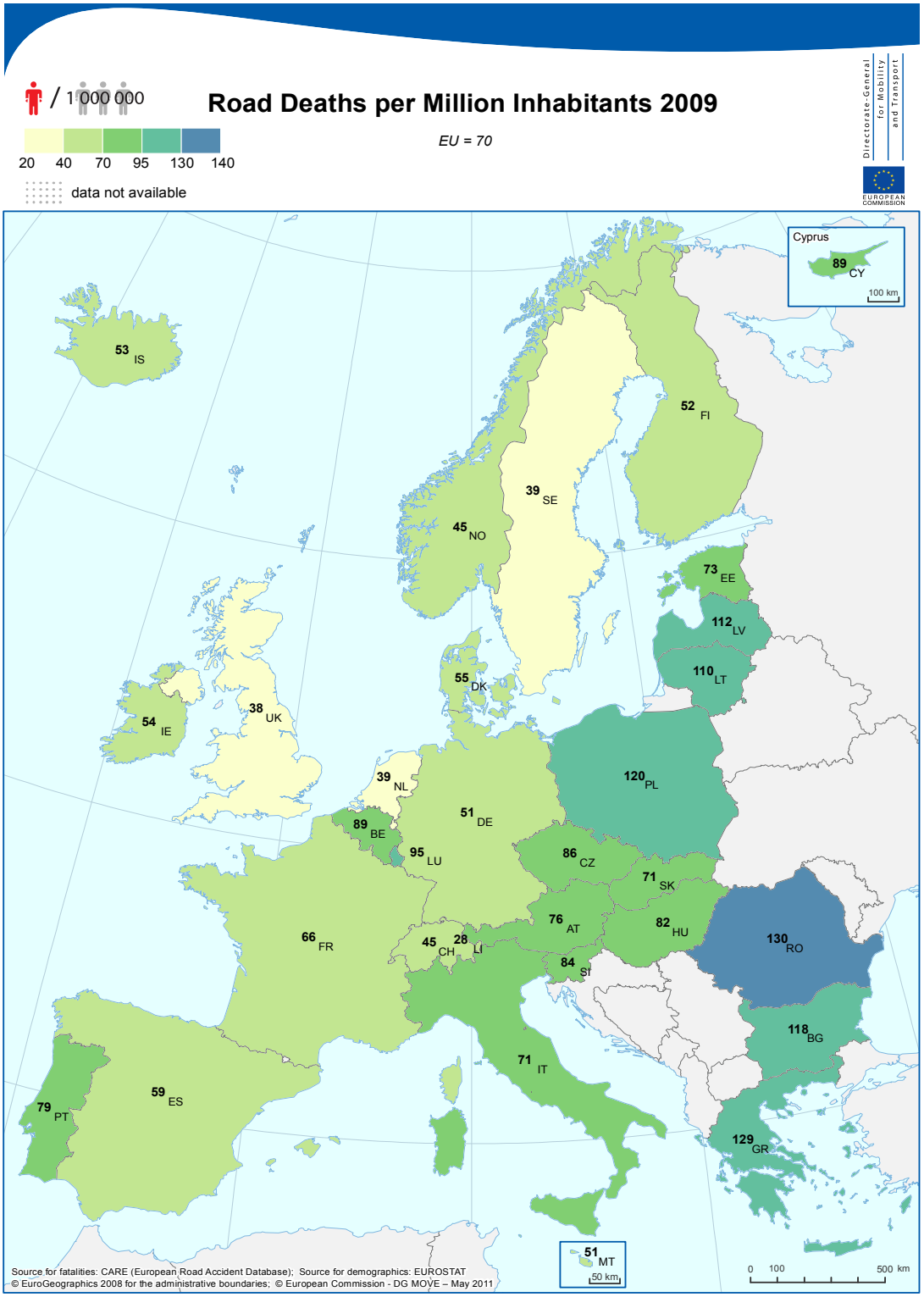


Fig. 3: Fallecidos en carretera por millón de habitantes (2009)
Tomado de CARE (EU road accidents database)

Todo esto implica que, en esta última década, más de 102.000 vidas se han salvado gracias a las medidas que se han ido implantando en los diferentes países de la Unión, persiguiendo la consecución del ambicioso objetivo de reducción del 50% en el número de fallecidos. Esto representa un total de más de 100.000 personas que habrían fallecido, si el sistema de tráfico hubiera seguido asumiendo

las cifras que se estaban dando al principio de la década, cuando los datos se situaban por encima de las 55.000 víctimas anuales en todo el territorio europeo.

Aún así, cerca de 31.000 personas perdieron su vida en las carreteras europeas en el último año del proceso. El dato en España es de 2.478 fallecidos en el año 2010 (DGT, 2011). y 2.060 en el año 2011 (DGT, 2012). La vía para continuar con la mejora y reducción de estos niveles permanece abierta, no está ni mucho menos agotada. Aún se puede hacer mucho trabajo y se pueden existen medidas que tomar para seguir reduciendo estas cifras.

En ese sentido, la Unión Europea ha fijado un nuevo objetivo para 2020 y de nuevo busca reducir en un 50% los datos actuales. Eso implicará reducir las cifras hasta los 15.000 fallecidos anuales. El punto de partida es objetivamente mejor que aquel del año 2001, pero, aún así, resulta sensiblemente mejorable.

1.3 Medidas de Seguridad Vial en España

Es erróneo pensar que la conceptualización del *accidente* como un evento aleatorio, del que no es posible precisar lugar ni momento de ocurrencia, lo convierte en algo *inevitable*. Indudablemente los accidentes tienen un componente importante de azar que determina, en parte, su ocurrencia; pero no es el único aspecto que actúa en todo ese proceso.

La existencia, o ausencia, de factores de riesgo aumenta, o disminuye, la probabilidad de sufrir un accidente. Eso explica, por ejemplo, por qué, a pesar de que el número de conductores de menos de 30 años no supera el 25% del total, en ese grupo se producen aproximadamente el 35% de las víctimas mortales en accidentes de tráfico; o por qué, proporcionalmente, se producen muchos más accidentes cuando se circula de noche, a pesar de que la densidad del tráfico es mucho menor. **El hecho de que sean impredecibles no hace que los accidentes sean inevitables**. En la introducción de su *Manual de medidas de Seguridad* Vial, Elvik y Vaa ejemplifican perfectamente esta línea de pensamiento (Elvik y Vaa, 2006).

En ese mismo libro los autores definen lo que se puede considerar como una *Medida de Seguridad Vial* aludiendo a cualquier instrumento técnico, programa o iniciativa que está orientada, ya sea de manera única o de forma explícita, a la mejora de la seguridad vial (Elvik y Vaa, 2006)

Tomando esta definición como referencia, en este apartado 1.3 se describen las principales medidas que fueron puestas en marcha en todo el proceso de mejora de la Seguridad Vial y de reducción del número de fallecidos en accidente de tráfico,

expuesto en el apartado anterior, que se dio a nivel europeo, y donde el caso español es arquetípico.

Obviamente lo complejo será establecer la relación causal entre las medidas y los diferentes efectos conseguidos. Para ello los estudios que se realicen deben dotarse de un estricto control metodológico. Lamentablemente esto no siempre se consigue en todas las ocasiones, por lo que el establecimiento de la validez de una medida de seguridad concreta puede adolecer de algún tipo de contaminación metodológica o sesgo. Este punto, aunque relevante, no es objeto principal de esta tesis. Pueden consultarse textos sobre esta temática en (Crossen, 1994; Hauer y Persaud, 1983).

En los últimos años la mejora de la Seguridad Vial en España ha sido muy significativa y se ha dado en un tiempo razonablemente corto. En el decenio 2001-2010 las muertes en accidente de tráfico descendieron en España un 55% (Tabla 2). Dentro del grupo de países de la Unión Europea este dato implica el cuarto mayor descenso, solo mejorado por los países bálticos (ver Fig. 2, p.12).

Total	2001	2006	2007	2008	2009	2010	2001/2010
Accidentes con víctimas	100.393	99.797	100.508	93.161	88.251	85.503	-15%
Fallecidos	5.517	4.104	3.823	3.100	2.714	2.478	-55%
Heridos Graves	26.566	21.382	19.295	16.488	13.923	11.995	-55%
Heridos Leves	123.033	122.068	123.226	114.459	111.043	108.350	-12%
Fallecidos por 100 accidentes	5,5	4,1	3,8	3,3	3,1	2,9	

Tabla. 2: Evolución de las cifras de accidentalidad en España 2001-2010 Fuente :DGT

Este descenso de las cifras de accidentalidad se produce en un contexto en el que todos los indicadores de exposición aumentan (DGT, 2011). El parque móvil se elevó, durante ese mismo periodo, un 33%. Lo mismo sucede con el censo total de conductores; de 2001 a 2010 hay un 20% más de personas con permiso para circular por las carreteras españolas. En cifras globales el volumen de tráfico aumentó un 16% en esos diez años.

Observando la evolución histórica de las cifras de muertos en carretera se constata, como se comentó anteriormente, que los datos presentan un periodo

alcista desde mediados hasta finales de los ochenta (Fig.4). Así mismo, se puede ver cómo posteriormente se produjo un descenso, a principios de los noventa, que se estancó en el año 1995, abriendo paso a un periodo, también reseñado anteriormente, sin una tendencia clara, hasta 2004. En ese momento, la Seguridad Vial se convierte en una de las prioridades de actuación del Gobierno Español.

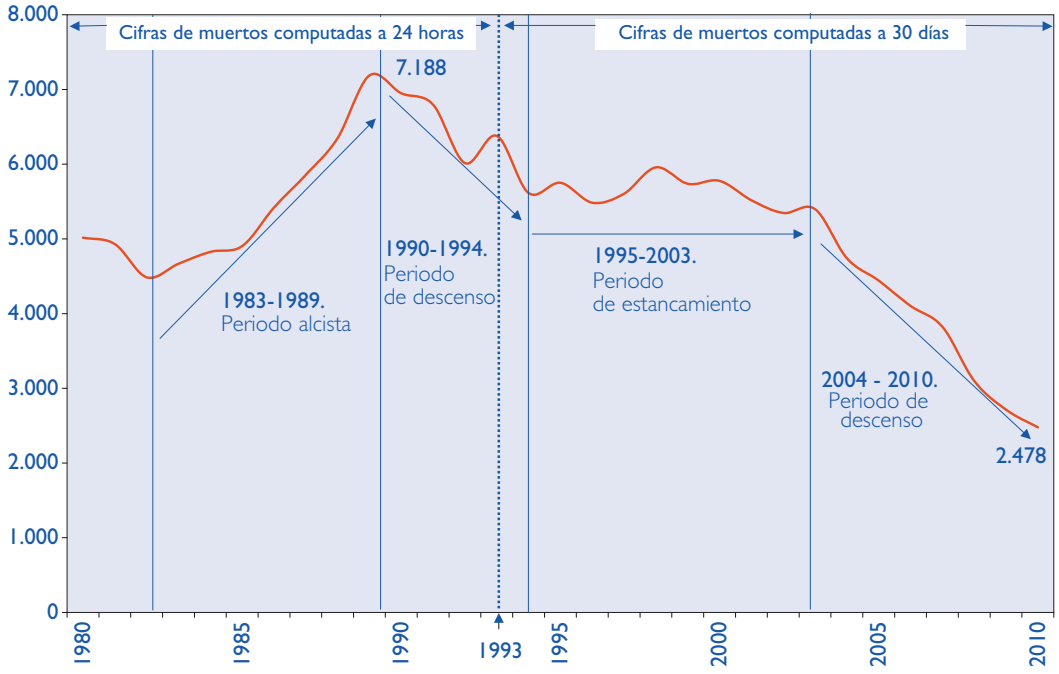


Fig. 4: Evolución del número de fallecidos en accidente de tráfico 1980-2010
Tomado de DGT Informe de siniestralidad 2010

La creación en el Parlamento de un Comité de Seguridad Vial y el inicio, en 2004, de los trabajos bajo la supervisión de la Dirección General de Tráfico, del Observatorio Nacional de Seguridad Vial, marcan el punto de partida de una serie de medidas encaminadas a mejorar los datos de siniestralidad y reducirlos de manera significativa (DGT, 2011).

Ese mismo año, se activa el Plan Estratégico de Seguridad Vial 2005-2008. Por primera vez el planteamiento de la estrategia se hace de manera plurianual, con una visión y objetivo claros: la reducción en un 40% del número de fallecidos en carretera.

Para ello se ponen en marcha 182 actuaciones, enclavadas en 10 áreas de trabajo en las que se monitorizan 26 indicadores (DGT, 2011).

De dichas medidas cabe destacar las siguientes:

1. **Implantación del Sistema de Permiso por Puntos**. Usando la posibilidad de pérdida del permiso por acumulación de infracciones como herramienta para corregir los malos hábitos al volante.

2. **Refuerzo de los medios humanos y técnicos**. Llegando, en 2010, a los 10.000 efectivos en la Agrupación de Tráfico de la Guardia Civil e incrementando el número de radares, en un 63% los móviles y en un 100% los fijos.

3. **Reforma del Código Penal y creación de la Fiscalía de Seguridad Vial**. Poniendo el foco en la acción penal sobre los infractores y los reincidentes.

Sin lugar a dudas, **la introducción del Sistema de Permiso por Puntos** ha sido el elemento estrella de este Plan. Un año después de su introducción en 2006, los índices de uso del cinturón se habían incrementado en un **10%** para los cinturones delanteros (de un 83% a un 93%) y en un **18%** para el uso de los traseros (de un 52% a un 70%) (DGT, 2011). Este aumento ha continuado a lo largo de los siguientes cinco años llegando a alcanzar el **92%** de uso en las plazas delanteras y el **85%** en las traseras, para desplazamientos en trayectos por carretera (Fig. 5). Para desplazamientos en zona urbana los datos aún son sensiblemente inferiores: un **86%** de uso en las plazas delanteras y un **68%** en las traseras. Parece como si, aún hoy, permaneciera la creencia de que en los desplazamientos urbanos, por su inferior velocidad, la ausencia de uso del cinturón tiene unos riesgos menores.

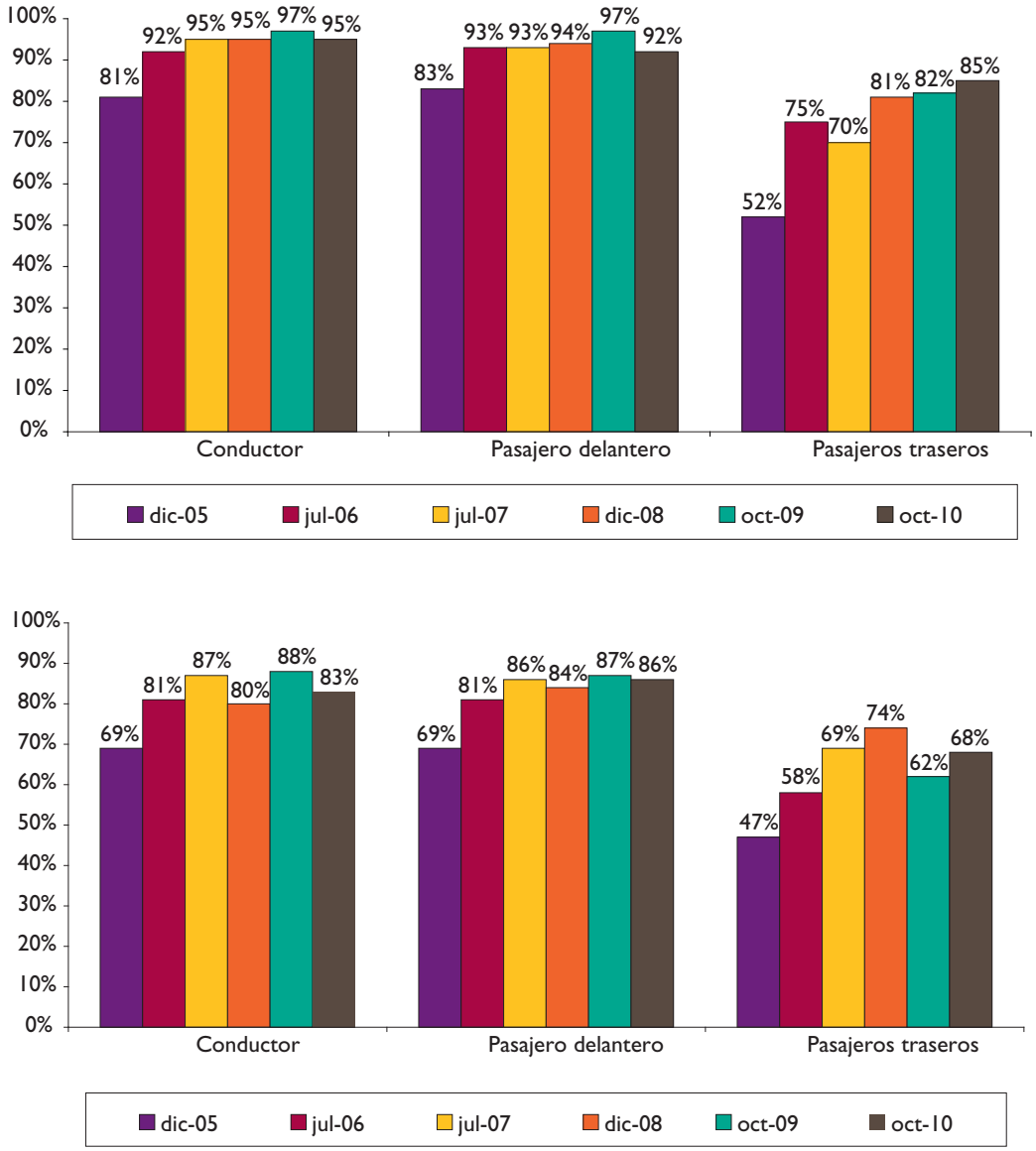


Fig. 5: Porcentaje de uso de cinturón de seguridad en turismos y furgonetas CARRETERA (arriba)/ ZONA URBANA (abajo).
Fuente: (DGT, 2011)

La segunda medida significativa ha sido **el refuerzo de medios, tanto humanos como técnicos**, dedicados a la vigilancia y control de la Seguridad Vial en las carreteras españolas. Esto ha permitido, por ejemplo, elevar el número de controles de velocidad (DGT, 2011).

En el año 2010 se determinó la existencia de velocidad inadecuada en un **13%** de los accidentes, lo que representa un descenso moderado comparado con el 15% del año 2001. Sin embargo, sí se aprecia un aumento del respeto a los límites de velocidad. A medida que ha ido creciendo el número de controles de velocidad y el de vehículos controlados, ha ido descendiendo el número de vehículos

denunciados (DGT, 2011). Un total de 684.385 vehículos denunciados en 26 millones y medio de controles durante el año 2010 (Fig. 6).

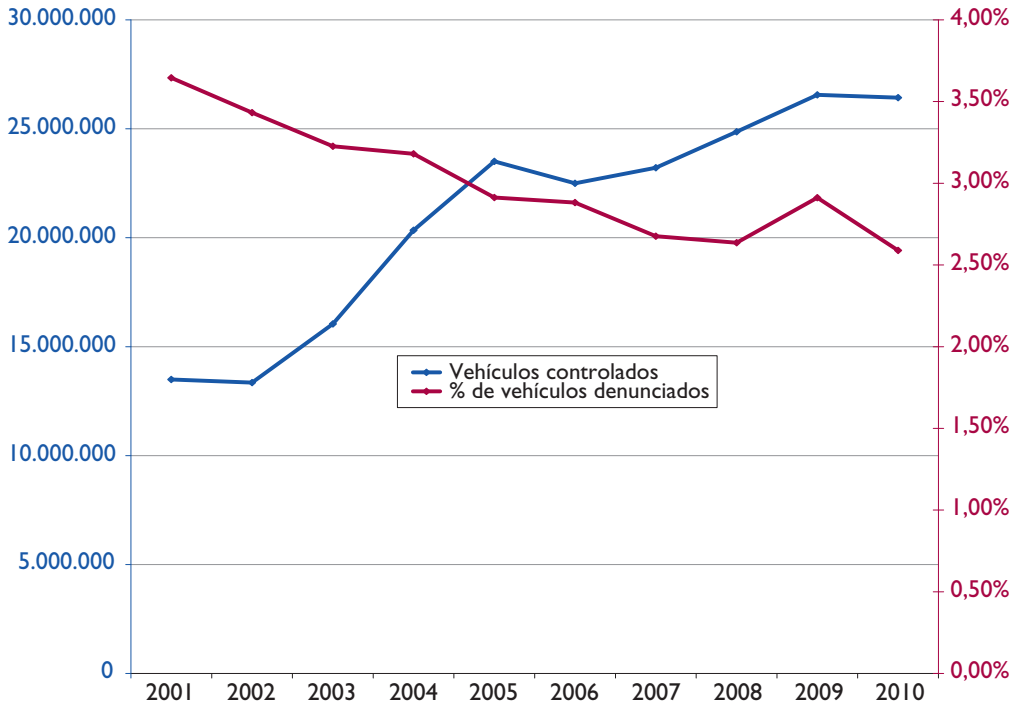


Fig. 6: Controles de velocidad realizados por la Agrupación de Tráfico de la Guardia Civil
Fuente: (DGT, 2011)

En último lugar, **la reforma del Código Penal** supuso que determinadas infracciones de tráfico cobraran peso como infracciones penales, siendo por ello motivo de cárcel para quien las comete. Entre ellas se encuentran la conducción temeraria o hacerlo careciendo del permiso requerido; también superar la velocidad máxima en 60 km/h en áreas urbanas o en 80 km/h en carreteras y autopistas. Por último, niveles de alcoholemia superiores a 1,2 g/l o a 0,6 g/l para conductores profesionales o conductores noveles, suponen también el tratamiento de la infracción como un delito penal del que se hace cargo la Fiscalía de Seguridad Vial.

En el periodo de 2001 a 2010 se ha logrado reducir en un 50% el número de positivos en controles toxicológicos de alcoholemia. Al igual que sucede con los controles de velocidad, a medida que ha aumentado el número de pruebas y controles preventivos ha disminuido el porcentaje de positivos. En el año 2010 solamente el 1,8% de las 4.550.000 pruebas realizadas dieron como resultado una tasa de alcoholemia positiva (Fig. 7).

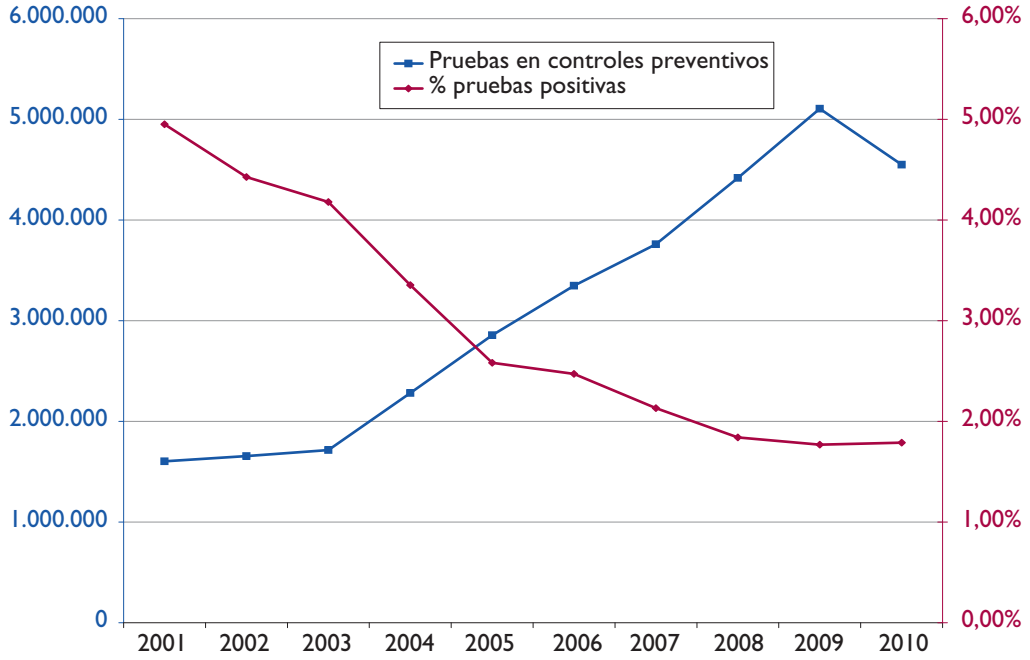


Fig. 7: Controles de alcoholemia realizados por la Agrupación de Tráfico de la Guardia Civil. Fuente: (DGT, 2011)

Todos estos pasos dados para mejorar la seguridad en carretera y minorar los efectos de los posibles accidentes se han tratado aquí, hasta este momento, desde una perspectiva global. A continuación se planteará la cuestión desde la óptica del conductor profesional, su implicación en la Seguridad Vial y la relación del sector de transporte de viajeros por carretera con las medidas y controles específicos que contribuyen a la mejora de la Seguridad Vial.

1.4 La siniestralidad en autobús y el sector de Transporte de Viajeros por Carretera

El autobús es un vehículo automóvil concebido y construido para el transporte de personas, con capacidad para más de nueve plazas, incluido el conductor (RDL 339/1990 vigente a 25 de mayo de 2010). La norma diferencia también, atendiendo al uso al que se destina, entre *Autobús* y *Autocar*, siendo el primero utilizado para desplazamientos de corto recorrido, con paradas frecuentes, con mayor facilidad de acceso, piso bajo y que dispone de plazas habilitadas para viajar de pie. El término *Autocar* se reserva para aquellos vehículos acondicionados específicamente para viajes interurbanos de largo recorrido en los que no se autorizan plazas de pie. En la terminología habitual no se diferencia un tipo de vehículo del otro y se hace referencia, de manera genérica, al término *Autobús* para referirse indistintamente a ambos.

Sin embargo, la Directiva 70/156/CEE establece una clasificación clara que categoriza los vehículos de acuerdo con su capacidad de asientos, uso principal al que se destina y masa máxima autorizada. Es lo que se conoce como *Definición M*.

En esta clasificación se estructuran los vehículos destinados al transporte de viajeros de la siguiente forma:

- Categoría M1: Aquellos vehículos de transporte de viajeros con no más de 8 plazas autorizadas, sin incluir el asiento del conductor.

- Categoría M2: Vehículos con más de 8 plazas autorizadas, sin incluir el asiento del conductor, y con una Masa Máxima Autorizada (MMA) inferior a 5 toneladas.

- Categoría M3: Engloba a los automóviles destinados al transporte de viajeros con más de 8 plazas autorizadas, sin incluir el asiento del conductor, y con una Masa Máxima Autorizada superior a 5 toneladas.

A su vez, esta clasificación se subdivide en tres clases I, II y III. La definición de *Autobús* se corresponde con vehículos de las categorías M2 y M3 pertenecientes a la clase I, es decir con áreas expresamente reservadas para pasajeros que viajen de pie.

La definición reservada para *Autocar* se aplica a vehículos de las categorías M2 y M3 pertenecientes a las clases II y III, siendo los vehículos de clase II aquellos que, mayoritariamente están construidos con plazas sentadas, pero reservan un espacio autorizado para plazas de pie. Los vehículos de la clase III no admiten la homologación de plazas para viajeros de pie.

No obstante, es habitual diferenciar las estadísticas, de acuerdo con el tipo de recorrido que realizan, en urbano o interurbano. Este último hecho no implica que los recorridos urbanos los realicen únicamente autobuses de acuerdo con la definición estricta aquí reseñada. Por ejemplo, una gran cantidad de los recorridos de transporte escolar son exclusivamente urbanos, pero se realizan en autocares acondicionados para viajes de largo recorrido.

El autobús es, además, uno de los medios, de transporte terrestre, más seguros que existen. Los accidentes con víctimas en los que están implicados autobuses representan un porcentaje muy bajo de los accidentes que se dan en la UE y en España. En la figura 8 se recoge el dato porcentual de fallecidos en accidente de tráfico conforme el tipo de vehículo.

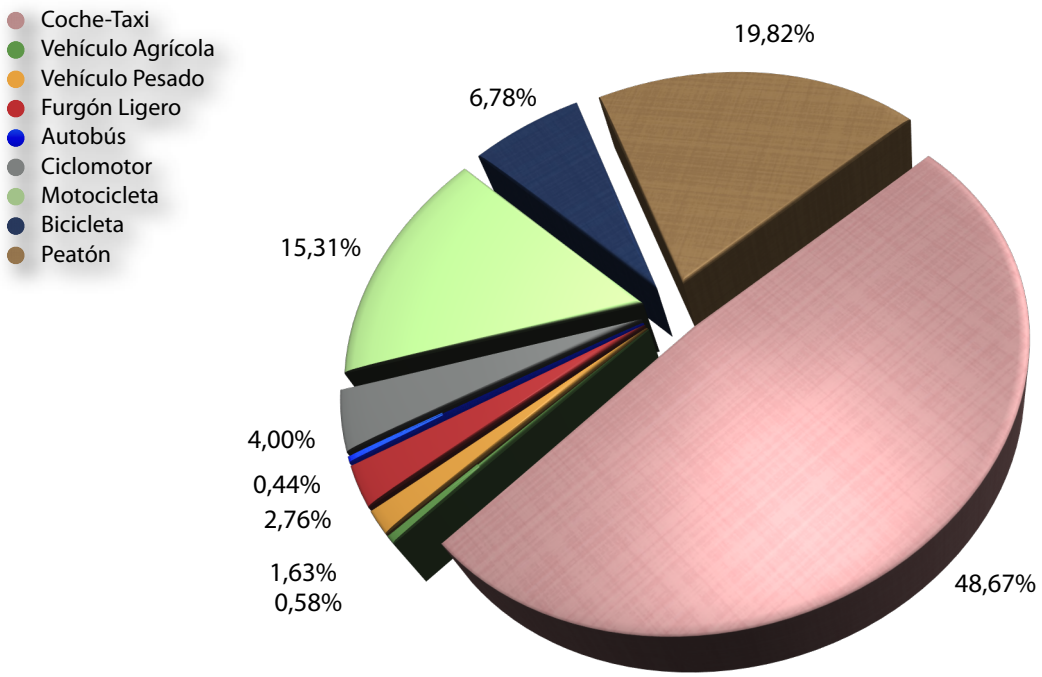


Fig. 8: Distribución de fallecidos por tipo de vehículo. (UE 2009)
Fuente CARE (EU road accidents database)

Estos datos conforman una estadística, del año 2009, de todos los países que forman parte de la base de datos CARE. Se puede observar como el porcentaje de muertos en autobús (0,44%) es el más bajo de todos, siendo incluso inferior al número de los que se produjeron en vehículos agrícolas (0,58%). En ese año en la UE fallecieron 148 personas en accidentes de autobús, 18 de ellos en España, siendo 17 pasajeros y 1 conductor.

La cifra de fallecidos en España en este medio de transporte, en la pasada década, varía y no muestra una tendencia clara. La media anual de fallecidos en autobús desde el año 2001 es ligeramente superior a 22 personas, con un máximo de 50 fallecidos en 2001, frente a un mínimo de cuatro fallecidos en los años 2004 y 2010. Esto se puede explicar teniendo en cuenta que en un medio de transporte con un índice de riesgo (número de fallecidos según la movilidad) bajo y un número tan reducido de accidentes con fallecidos, un único accidente grave puede alterar la tendencia y las comparaciones entre años. En el año 2011, la cifra de fallecidos ha sido menor que en 2010, registrándose tres muertes (DGT, 2012)

La accidentalidad de los autobuses es sensiblemente superior en zona urbana que en carretera (Tabla 3). En 2010, el 80% de los accidentes de autobús se produjeron en áreas de circulación urbana. Sin embargo, se aprecian lesiones más graves en los accidentes que suceden en carretera.

	Acc. con víctimas	%	Muertos totales	%	Muertos ocupante	%	Muertos terceros	%	(*)
Carretera	353	20%	28	56%	2	50%	26	57%	93%
Zona urbana	1.424	80%	22	44%	2	50%	20	43%	91%
Total	**1.777**	**100%**	**50**	**100%**	**4**	**100%**	**46**	**100%**	**92%**

Tabla. 3: Accidentalidad de autobuses en España por área en que se producen, año 2010
Fuente DGT (elaboración propia)
(*) Porcentaje de muertos terceros sobre el total de muertos en accidentes con al menos un autobús implicado

Es de especial relevancia la última columna, donde se muestra el porcentaje de Muertos terceros sobre el total de muertos en accidentes en los que está implicado al menos un autobús. Se puede observar que, en la mayoría de los casos, las víctimas que salen peor paradas son las que viajan en los vehículos con los que colisiona el autobús. Para los países de la OCDE se ha estimado que el riesgo de fallecer o sufrir heridas graves en un accidente es entre siete y nueve veces menor para los pasajeros de un autobús, que para los pasajeros de un turismo (Albertsson y Falkmer, 2005).

Estos mismos autores trazan un perfil del accidente de autobús tipo en Europa. Su investigación concluye que la mayor parte de los accidentes se producen en zona urbana y en condiciones de seco. En la mayoría de las ocasiones, el accidente que se produce es la colisión, principalmente frontal o lateral, con otro vehículo; con mayor frecuencia contra turismos. Esto se puede ver también en García, Martin y Páez (2006). Por último, determinan que el tipo de accidente más relacionado con las consecuencias más graves es el vuelco. En estos accidentes, la intrusión de la estructura en el espacio de los viajeros o la proyección de los mismos son los mecanismos más frecuentes y graves, a la hora de producir lesiones.

Es por esto que distintas Directivas Europeas, incorporadas en los últimos años en materia de seguridad pasiva en autobuses y que son de obligado cumplimiento, han ido encaminadas a reducir las consecuencias lesivas de este tipo de accidentes. En concreto:

Para impactos frontales

- **Directiva 74/408-2005/39/CE**, sobre la resistencia de asientos y anclajes.

- **Directiva 76/115-2005/41/CE**, sobre los anclajes del cinturón de seguridad.

- **Directiva 77/541-2005/3/CE**, sobre cinturones de seguridad, sistemas de retención e instalación.

- **Directiva 2001/85/CE**, en lo relativo a sistemas de retención de sillas de ruedas y sus ocupantes.

Para vuelco

- **Directiva 2001/85/CE**, que incorpora como requisito el ensayo de estabilidad de vuelco, exigiendo que los vehículos soporten, sin volcar, una inclinación lateral de 28 grados sobre plataforma.

- **Reglamento nº66**, que busca garantizar la resistencia de la estructura en caso de vuelco, preservando un espacio de supervivencia para los pasajeros.

Sobre este particular puede encontrarse una excelente descripción de los objetivos de los diferentes reglamentos y directivas en Vicente, Martín, Grimaldi y Aparicio (2008).

Así las cosas, durante el año 2009, en España, se produjeron 929 víctimas en 353 accidentes de tráfico en carretera con al menos un autobús implicado, disgregadas de la siguiente forma: **40** fallecidos (18 ocupantes y 22 terceros), **140** heridos graves (77 ocupantes y 63 terceros) y **749** heridos leves (431 ocupantes y 318 terceros (DGT, 2011).

El perfil del accidente tipo se ajusta bastante al descrito anteriormente. El siniestro más frecuente fue la *colisión entre vehículos*, que supera el **78%** (78,47%). El segundo tipo de accidente, *salida de la calzada*, no llega al **10%** del total (9,35%).

El tipo de vía en el que se producen con mayor frecuencia es la *carretera convencional* (**68,27%**), seguido por la *autovía* (**24,08%**).

En cuanto a los factores concurrentes en el accidente, los que mayor presencia tienen son la *infracción a norma de circulación* y la *distracción*. Es significativo que, en ninguno de los siniestros documentados en 2009, la *condición*

de la señalización, los *tramos en obras* o el *mal estado del vehículo* constituyeron factor concurrente en la ocurrencia del accidente.

Por último, en este perfil, por lo que se refiere al conductor de autobús, en la mayoría de los accidentes se trata de un profesional por cuenta ajena, español, entre 30 y 49 años, con una antigüedad de permiso superior a 10 años y que sufre el accidente durante su jornada laboral.

Se da la circunstancia de que los conductores de autobús, en el ejercicio de su profesión, poseen una serie de características particulares que los constituyen como un grupo diferente al resto de los conductores, principalmente de turismo, con los que comparten el entorno de la circulación (Dorn y af-Wåhlberg, 2008).

Más adelante se tratará la cuestión relacionada con el factor humano en la conducción y su relación con la Seguridad Vial. Baste, de momento, con apuntar aquí esas características que rodean a la profesión de conductor de autobús y que se relacionan directamente con su riesgo de sufrir un accidente.

1. Los conductores de autobús poseen **experiencia previa con turismo** antes de acceder al permiso de la clase D que les habilita para conducir autobuses. Al contrario de lo que pudiera parecer, esa experiencia previa **no les beneficia**. De hecho, no se han encontrado asociaciones significativas entre la antigüedad del permiso y los accidentes de autobús, (af-Wåhlberg, 2005).

2. **En un entorno complejo de tráfico**, con un vehículo de grandes dimensiones y manejo complejo, tienen la **presión añadida** de la responsabilidad **sobre las vidas de los pasajeros**.

3. Su **kilometraje anual** está muy por encima del de un conductor medio; las **largas jornadas de conducción aumentan** sensiblemente **su exposición al riesgo**.

4. Por último, se añaden los efectos negativos de diferentes **factores organizacionales** (presión por cumplimiento de horarios, tiempos de conducción, nocturnidad, trabajo a turnos, etc.)

Es por estos motivos por los que la profesión de conductor de autobús está fuertemente regulada y sometida a una legislación muy exigente, orientada a garantizar que el trabajo se realiza en condiciones propicias para la Seguridad Vial.

En primer lugar, desde los años sesenta se ha legislado el **acceso a la profesión**, regulando, con diferentes niveles de restricción, las condiciones para la obtención del permiso de la clase D, que es el que en España autoriza para conducir vehículos de transporte de viajeros de las categorías M2 y M3. Desde la exigencia obligatoria de acreditar, como mínimo, un año de experiencia profesional con un vehículo pesado de transporte de mercancías para poder examinarse y obtener el permiso de autobús; pasando por la condición restrictiva, con código armonizado 101 (que limitaba el radio de acción de un conductor de autobús con menos de un año de experiencia con vehículos de más de 3'5 toneladas a 50 kilómetros), se ha llegado a la situación actual, en la que el **R.D. 818/2009,** de 8 de mayo, limita la edad mínima para la obtención de este permiso a los 24 años cumplidos.

La edad resulta, de esta forma, un primer elemento de control. La idea es limitar el acceso hasta que se alcance una edad que presupone una cierta madurez, un bagaje práctico al volante, así como un mínimo conocimiento del entorno del tráfico. Edad y experiencia son dos variables tradicionalmente estudiadas en su relación con la accidentalidad y el riesgo a sufrir un accidente (Clarke, Ward, y Jones, 1998; Mayhew, Simpson, y Pak, 2003; McGwin y Brown, 1999; Ryan, Legge, y Rosman, 1998), aunque casi siempre para conductores de turismo. Se ha estudiado poco su relación con conductores profesionales y si esos resultados pueden ser generalizables.

La tabla 4 muestra las condiciones de obtención que están actualmente en vigor en España.

Clase de permiso	Vehículos que habilita a conducir	Edad mínima
D1	Vehículos de transporte de viajeros hasta 16 plazas (excluido el conductor) y hasta 8 metros.	21 años cumplidos
D	Vehículos de transporte de viajeros de más de 8 plazas (excluido el conductor).	24 años cumplidos
D1+E	Conjunto de vehículos con automóvil clase D1 y remolque de más de 3'5 t MMA	21 años cumplidos
D+E	Conjunto de vehículos con automóvil clase D y remolque de más de 3'5 t MMA	24 años cumplidos

Tabla. 4: Condiciones para la obtención de los permisos de autobús. R.D. 818/2009 (Elaboración propia)

El segundo aspecto que regula el acceso a la profesión es la **Formación**. Tradicionalmente no existe un nivel formativo de base requerido para la obtención del permiso de la clase D. En el año 1976 la **Directiva 76/914/CEE** establece una formación mínima obligatoria para los conductores de vehículos pesados, pero sin que adopte la forma de un título oficial reglado. En ese primer intento de regular la formación de los conductores profesionales, se establecían como obligatorios unos conocimientos básicos sobre:

- La construcción del vehículo y el uso de sus componentes principales.
- El sector de transporte y la administración de las empresas.

Se exigía además acreditar experiencia en la conducción de vehículos de transporte de mercancías o de viajeros.

Posteriormente, se publicó el **Real Decreto 1998/1996**, de 6 de septiembre, por el que se regulaba el Certificado de Profesionalidad de la ocupación de Conductor de Autobús. En este caso, que sigue vigente y aún se imparte, la formación, la cual no se establece como obligatoria, consiste en 405 horas lectivas, encaminadas a la mejora y perfeccionamiento de la conducción, así como a la transmisión de conocimientos básicos para el ejercicio de la profesión como son: la atención al público, el manejo de máquinas de cobro o el mantenimiento mecánico general del vehículo.

Actualmente, la **Directiva 2003/59/CE** vino a intervenir en el acceso a la profesión, atendiendo a los aspectos formativos.

Esta norma europea regula la cualificación inicial obligatoria para determinados conductores. En este caso concreto, la directiva afecta a los conductores de vehículos de transporte de viajeros, para los que es requisito necesario estar en posesión del permiso de conducir de la clase D. El texto busca el cumplimiento de dos objetivos:

- La mejora de la Seguridad Vial y de la seguridad del conductor, favoreciendo actitudes seguras hacia la conducción y dando pautas claras de actuación ante accidentes.

- Facilitar que los conductores puedan responder a las nuevas exigencias originadas por la evolución del mercado.

En España la transposición de la directiva se realizó mediante el **Real Decreto 1032/2007**, de 20 de julio, por el que se regula la cualificación inicial y continua de los conductores de determinados vehículos destinados al transporte por carretera. En este R.D. se establece la formación inicial como un requisito obligatorio para todos aquellos conductores de autobús que deseen ejercer su profesión realizando servicios de transporte público de viajeros y hayan obtenido su permiso de conducir, de la clase correspondiente, después del 10 de septiembre de 2008.

También se contempla esta formación como un proceso distinto y necesariamente diferenciado de la obtención del permiso. La formación en este Certificado de Aptitud Profesional (CAP), se convierte de esta forma en un complemento necesario del permiso de conducir. Así se articula que, para poder ejercer la profesión, necesariamente habrá de acreditar ambos certificados: por un lado, el permiso de conducir de la clase requerida para conducir y por otro, el Certificado de Aptitud Profesional.

Para obtener este certificado, el conductor deberá superar, en un centro autorizado, un curso de formación en alguna de estas dos modalidades:

- **Ordinaria**: 280 horas lectivas, de las cuales 20 han de ser de conducción individual.

- **Acelerada**: 140 horas lectivas de las cuales 10 han de ser de conducción individual.

Tras la realización del curso de formación el alumno tendrá que superar un examen, convocado por el Órgano correspondiente de su Comunidad Autónoma, para obtener la tarjeta acreditativa oportuna (Fig. 9).

La expedición del certificado se anota en el Registro General de Transportistas y Empresas de Actividades Auxiliares y Complementarias del Transporte. Se realiza sin plazo de validez, pero sujeta a la validez de la tarjeta, que es de cinco años. Antes de concluir este periodo de cinco años el conductor habrá de realizar un curso de formación continua, presencial y con una duración de 35 horas. La superación de este curso, exento de realización de examen, supone la renovación automática de la vigencia de la tarjeta CAP por otro periodo de cinco años.

Fig. 9: Anverso y reverso de tarjeta CAP

Además de las obligaciones establecidas para el acceso a la profesión que, como puede observarse, tienen un claro carácter limitativo, el Legislador ha establecido todo un marco normativo que regula la profesión, en lo que se refiere a condiciones de trabajo, con el objetivo de favorecer la Seguridad Vial.

Más allá de lo dispuesto en el Estatuto de los Trabajadores, teniendo en cuenta el especial carácter de esta profesión, el **R.D. 1561/1995** de 21 de septiembre, sobre jornadas especiales de trabajo, regula en sus artículos 10 a 12 el tiempo de trabajo en los transportes por carretera. De esta forma, en España, el tiempo de trabajo efectivo para un conductor no podrá superar las 48 horas semanales de promedio en cómputo de cuatro meses, ni exceder en ningún caso las 60 horas semanales.

Así mismo, el **Reglamento (CE) 561/2006** relativo a la armonización de determinadas disposiciones en materia social en el sector de los transportes por

carretera, establece los periodos máximos de conducción y mínimos de descanso, tanto diarios como semanales, de obligado cumplimiento para un conductor.

El control del cumplimiento de esta norma corre a cargo del **Reglamento (CEE) 3821/85,** relativo al aparato de control en el sector de los transportes por carretera. En él se establece la norma por la que se regula la instalación, homologación y control del llamado Tacógrafo. Este aparato, de instalación obligatoria en todos los autobuses destinados al servicio público de transporte de viajeros, es un dispositivo de registro analógico que, mediante una serie de punzones, deja impreso, sobre una hoja preparada a tal efecto (Fig. 10), registro de la distancia recorrida, velocidad, tiempo de trabajo o descanso y las aperturas realizadas en el aparato.

Actualmente, el **Reglamento (CE) 2135/98**, por el que se modifica el Reglamento (CEE) 3821/85 relativo al aparato de control en el sector de los transportes por carretera, regula la instalación y uso de un Tacógrafo Digital, que mejora las prestaciones del anterior, en cuanto a capacidad de registro y control así como resistencia ante el fraude. En este caso, el registro se produce en una tarjeta, expedida específicamente para cada conductor, que es personal e intransferible y que detalla su actividad de los últimos 28 días, (Fig. 10).

Fig. 10: Hoja de registro analógico y tarjeta de tacógrafo digital

Todo este cuerpo normativo se da en un sector económico que está aún muy atomizado, aunque existe una tendencia a la concentración empresarial, fuertemente intervenido por la administración y que tradicionalmente adolece de la ausencia de sistemas de gestión de trabajadores, especialmente en cuanto a reclutamiento y selección, profesionalizados y objetivos.

A finales de 2010 existían en España 4.533 empresas de transporte de viajeros en autobús para un total de flota de 40.768 vehículos, con una antigüedad media de 11 años (DGTT, 2011b). De ellas, 35 explotan algunas de las 101 concesiones de transporte regular titularidad de la Administración del Estado. El resto, se reparte entre concesiones regulares de las Comunidades Autónomas y las autorizaciones para realizar transporte regular de uso especial y transporte discrecional. En 1999 la flota media en las empresas de transporte de viajeros era de 7'2 vehículos; actualmente, se sitúa en torno a 10'6 (DGTT, 2011a), lo que refleja el proceso de concentración empresarial que se está produciendo. Aún así, en estos momentos la característica del sector continúa siendo que las empresas sean pequeñas, con un número de vehículos no superior a cinco y con un estilo de gestión de empresa familiar.

En esta situación, a pesar de los esfuerzos normativos por limitar y controlar el acceso a la profesión, poco se ha hecho por la profesionalización de los procesos selectivos para acceder a un puesto de trabajo. En general, para realizar la selección de un nuevo conductor, se confía el criterio de la experiencia previa y en la superación de las pruebas correspondientes al reconocimiento psicotécnico.

Todas las medidas descritas hasta ahora han ido orientadas hacia la mejora de la Seguridad Vial, pero sin hacer mención explícita a uno de los criterios más relacionados con la siniestralidad en carretera y que es clave en el éxito futuro de un proceso selectivo, **el Factor Humano**.

Un error mínimo al principio puede ser máximo al final.

Aristóteles

Factor Humano y Selección de Conductores

2

Este capítulo se acerca a las teorías relacionadas con el comportamiento humano en la actividad de la conducción. Describe el recorrido histórico de las teorías, sobre factor humano, relacionadas con la accidentalidad y señala aquellas más significativas. Profundiza en el modelo de selección de conductores que se aplica en España y en cómo, a lo largo del siglo XX, ha ido estableciéndose como un modelo de referencia en la evaluación de las capacidades y aptitudes necesarias para una conducción segura.

Finalmente, se exponen algunas de las pruebas que actualmente se utilizan para la evaluación de conductores, se comentan sus limitaciones y se introducen los motivos por los que se detecta una necesidad de creación de una prueba como la que aquí se presenta.

2.1 Antecedentes históricos: Teorías y modelos explicativos

En el momento actual la concepción generalmente aceptada en los principales centros de investigación (National Highway Traffic Safety Administration, European Transport Safety Council o el Instituto de Investigación en Tráfico y Seguridad Vial) es que los accidentes no son eventos que se puedan explicar a partir de un único factor o motivo. Por el contrario, es una concatenación de diversas causas, todas diferentes, la que desencadena el accidente o aumenta la probabilidad de sufrirlo sin necesariamente llegar a padecerlo.

Así las cosas, los accidentes se tratan, como una realidad multicausal y se habla de factores de riesgo para dar explicaciones plausibles de por qué se producen. El actual Catedrático de Seguridad Vial de la Universidad de Valencia y miembro de la junta permanente del INTRAS (Instituto de Investigación en Tráfico y Seguridad Vial), el profesor don Luis Montoro, lo expresa así en su manual de Seguridad Vial del año 2000:

> [...] Los accidentes de tráfico no son, pues, el resultado de un factor simple, sino más bien el producto de una conjunción de muchos factores. (Montoro, Alonso, Esteban, y Toledo, 2000 pag. 33)

De esta forma, se entiende el sistema del tráfico como el resultado de la interacción de tres componentes generales donde se originan los principales factores de riesgo para la conducción. Entorno, Vehículo y Conductor (Fig. 11). Los trabajos de los últimos años han ido encaminados a determinar el peso específico

de cada uno de estos factores en la accidentalidad y en el riesgo de sufrir un accidente.

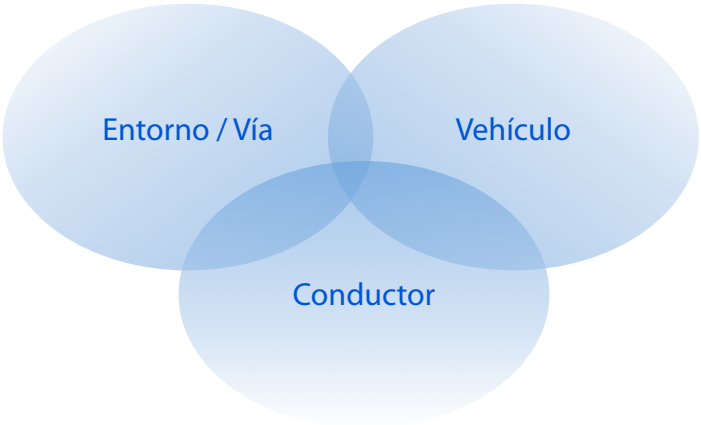


Fig. 11: Fuentes de factores de riesgo en la conducción

Los datos avalan la teoría de que, de entre todos los factores de riesgo para la conducción, son los relacionados con el conductor los que tienen mayor influencia en la accidentalidad. El llamado Factor Humano es, de una manera u otra, la principal causa de los accidentes de tráfico (Montoro, Lijarcio, Roca, y Puchades, 2007).

En la Fig. 12 pueden observarse los datos más actuales sobre las causas de accidentes, tanto en aquellos donde la causa estudiada es la principal como aquellos donde es la probable.

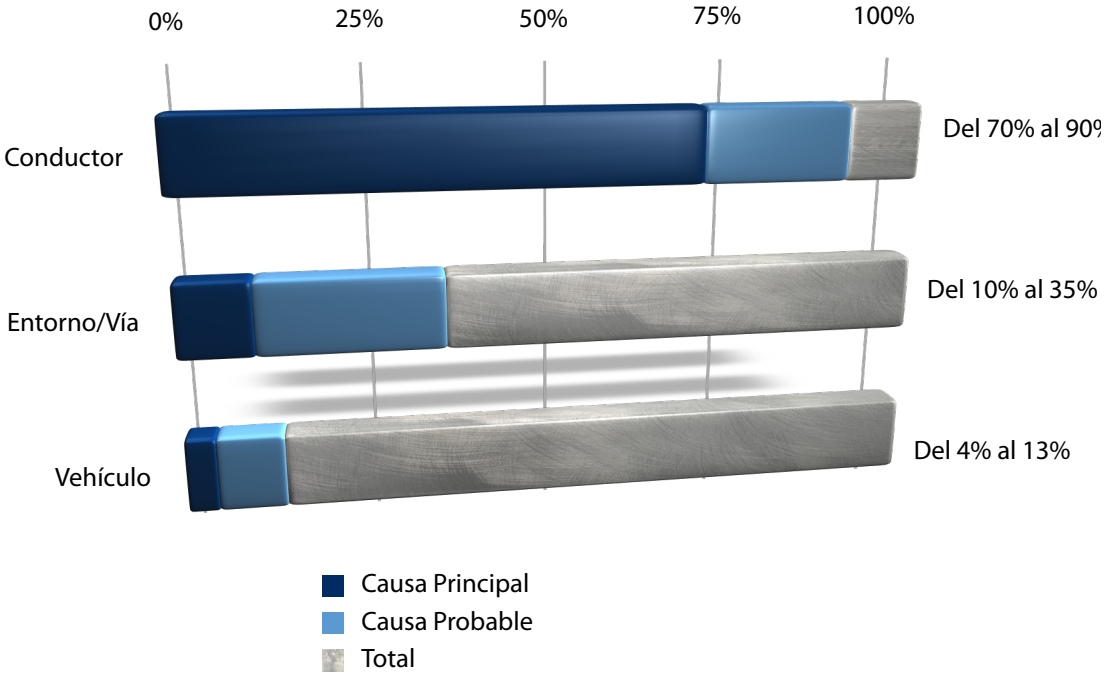


Fig. 12: Porcentaje de Implicación de las diferentes causas de los accidentes de tráfico. Fuente (Montoro, et al., 2007)

Como muestra el gráfico, entre un 70% y un 90% de los accidentes de tráfico se deben al Factor Humano. Esto implica necesariamente que en la gran mayoría de los accidentes existe una actuación humana errónea (Carbonell, Bañuls, Cortés, y Sáiz, 1995). De hecho, la percepción social, sin información específica sobre datos y estadísticas, es similar a la realidad de los datos. A principios de la década de los noventa más de un 84% de los conductores españoles consideraba que los accidentes se producían, principalmente, por un fallo del conductor, mientras que solo un 29% consideraba la implicación de un fallo mecánico como causa principal de los accidentes (del Río, Prada, Martín, y Alvarez, 1991).

En el caso de los autocares y el transporte de viajeros por carretera, esta relevancia del factor humano es incluso mayor. Diversos estudios, realizados con conductores de autobús, arrojan un 99% de influencia de la conducta o actuación del conductor en el accidente (af-Wåhlberg, 2004a, 2004b, 2005). Asimismo, un informe del Instituto Universitario de Investigación del Automóvil (INSIA) publicado en 2007 en uno de los boletines del Observatorio Nacional de Seguridad Vial, revela datos interesantes en torno a la accidentalidad de autobuses en España (DGT, 2007):

- Solamente el **7%** de los accidentes se producen por fallo mecánico.
- Cuando hay influencia de las condiciones de la infraestructura, lo que se da en un **31%** de los accidentes, la maniobra incorrecta del conductor es factor concurrente.
- El **31%** de los accidentes tienen origen en una pérdida de atención del conductor del autobús o del vehículo contrario.
- En un **15%** de los accidentes la causa está en una maniobra incorrecta de adelantamiento.

Sin embargo, esto no siempre fue así. Tratar de explicar por qué suceden los accidentes de tráfico y desentrañar sus causas es una tarea más compleja de lo que pudiera parecer en un principio. Incluso existen autores que han puesto en tela de juicio el uso semántico del término *causa* como explicación de los accidentes, negando su utilidad (Haight, 1980). En su lugar, definen los accidentes en términos de una compleja realidad aleatoria que, como tal, puede expresarse estadísticamente. De esta forma, lo importante sería centrarse en conocer aquellos factores que afectan al proceso estocástico, al menos aquellos que son conocidos, dando por hecho que muchos otros factores no podrán ser conocidos.

Aún así, desde hace un tiempo se ha venido poniendo el acento de la investigación sobre las causas de los accidentes y la participación del factor humano. Este recorrido histórico en el estudio de los accidentes y sus causas comienza a finales del siglo XIX, casi con la aparición del automóvil.

2.1.1 Modelos de habilidades

En 1898, hace poco más de cien años, se inicia la investigación de accidentes con la publicación del libro *La ley de los pequeños números* (Bortkiewicz, 1898). Su autor, Ladislaus Bortkiewicz, era en realidad un economista y sus intereses científicos no estaban orientados hacia los accidentes, pero uno de sus estudios más famosos tuvo una influencia enorme en la primera época de investigación sobre esta temática.

Bortkiewicz dedicó 20 años a recoger datos sobre los fallecidos por coz de caballos entre los soldados de los 14 cuerpos de caballería del ejército prusiano. El estudio de los datos de esos accidentes arrojó que la distribución real de los mismos se ajustaba perfectamente a la distribución de Poisson, que modelaba procesos aleatorios.

$$P(r)= \frac{e^{-\lambda}\lambda^{r}}{r!}$$

Donde *P(r)* es la probabilidad de un sujeto de sufrir *r* accidentes y λ es el número medio de accidentes.

Esto permitió deducir que los accidentes eran eventos aleatorios, impredecibles e incontrolables. Esta idea se aceptó de manera generalizada hasta que estudios realizados con la llegada de la Primera Guerra Mundial pusieron en tela de juicio esta primera visión.

Durante la Gran Guerra, con la mayoría de los hombres luchando en el frente, la incorporación de la mujer a los procesos de fabricación industrial fue un hecho. A esto se unió la incorporación de trabajadores más jóvenes y menos experimentados y unos tiempos de trabajo y jornadas con mucha mayor presión. Al final de la contienda, tras haber servido en una unidad de investigación médica, Major Greenwood, un médico inglés que había estudiado a principios de siglo con Karl Pearson, comenzó a interesarse por los problemas de salud y el incremento de

accidentes relacionados con el trabajo industrial. Junto con el estadístico Udny Yule, comprobó que existía una concentración de accidentes en grupos determinados de trabajadores, más allá de lo esperable por azar. Tras realizar diversos estudios durante 1919, al año siguiente publicaron un trabajo en el que, *por pura conveniencia práctica*, eligieron la curva binomial para representar la distribución de los datos (Greenwood y Yule, 1920). En este estudio determinaron que la distribución resultante, binomial negativa, se ajustaba a la realidad de los datos mejor que lo que lo hacía la distribución de Poisson.

$$P(x)=\binom{x-1}{x-k}q^{x-k}.p^{k}$$

Donde *p* es la probabilidad de obtener un resultado *k*; *q* es la probabilidad de no *p* y *x* es una variable aleatoria que recoge el número de pruebas necesarias para obtener *k* éxitos.

A pesar de los buenos resultados obtenidos y del aparente buen ajuste del modelo propuesto, Greenwood nunca llegó a afirmar, o asumir, una probabilidad inicial distinta de sufrir un accidente. Lo hizo de manera clara, sin dejar lugar a dudas.

> [...] Nuestra solución al problema de una diferenciación a priori [predisposición al accidente] fue empírica, en el sentido de que nuestra única justificación de la elección concreta de una curva de Pearson tipo III fue que su rango comenzaba en $\lambda=0$ y conducía a una forma estadísticamente útil. No sugerimos que hipotetizar una ley que conecte λ [predisposición] con *r* [número de accidentes] vaya a producir una distribución similar. Greenwood en (Cresswell y Froggatt, 1962).

Sin embargo, diferentes investigadores, basándose en este trabajo, dieron por probada esta probabilidad diferencial. Esto es, partir desde el inicio de una contribución al accidente diferente para cada sujeto. Farmer y Chambers bautizaron a este modelo como *Presdisposición al accidente*, asumiendo de esta forma que determinadas personas tenían más tendencia a padecer accidentes que otras (Farmer y Chambers, 1926). El siguiente paso natural era, por tanto, identificar a esas personas más proclives a sufrir accidentes.

Aprovechando el auge de la Psicología en esa época, se inició un estudio intensivo en la búsqueda de tests y pruebas psicológicas que permitieran detectar estos perfiles para poder intervenir sobre ellos, con el objetivo de actuar

preventivamente, evitando la aparición de accidentes provocados por estos sujetos con mayor tendencia a sufrirlos (Cresswell y Froggatt, 1962).

Este punto de vista, preponderante durante casi cuatro décadas (1920-1950), se tradujo en evaluaciones de destreza corporal, estudios psicológicos y medidas de las correlaciones entre variables implicadas; pero aún así no se encontró ninguna medida, prueba o batería de tests que permitiera discriminar a los muy propensos al accidente de aquellos que no lo eran. Era evidente que esta aproximación al problema no era capaz de aportar una explicación de los accidentes en su totalidad.

Aún así, toda esta línea de investigación ha dado como resultado una serie de modelos explicativos que tratan de contribuir a la mejora de la Seguridad Vial desde una perspectiva mecanicista en la que las **habilidades** del conductor para adaptarse a las exigencias de la circulación, carretera, entorno, vía y vehículo, son el elemento fundamental para la seguridad al volante (Carbonell, et al., 1995).

Este modelo ha llegado a nuestros días en forma de diversas formulaciones teóricas concretas; la más conocida es, la de Rockwell (Rockwell, 1981). En este planteamiento de habilidades, la conducción es entendida como una tarea de *pasos impuestos,* con una serie de exigencias, externas al conductor, a las que debe responder. Las habilidades perceptivas y motoras juegan entonces un papel preponderante. No en vano, Rockwell ha sido uno de los pioneros en el estudio de la relevancia de la visión y las capacidades visuales en la conducción (Carbonell, et al., 1995).

De esta forma, el conductor toma la información del tráfico utilizando sus canales perceptivos, principalmente el visual, y, tras procesarla, toma una decisión que se traduce en una conducta concreta. En esta relación *percepción - decisión* no hay demasiado espacio para los factores personales. El conductor se convierte de esta forma en un mero ejecutor de respuestas a pasos impuestos.

Con todas sus limitaciones, el planteamiento teórico de la *Predisposición al accidente* ha contribuido de manera significativa a la mejora de la Seguridad Vial en carretera (Carbonell, et al., 1995).

En primer lugar, **puso el acento en el Factor Humano** ya que, sin hacer referencia explícita a los factores de personalidad, se plantea un modelo en el que las capacidades humanas están detrás de las causas de los accidentes. Esto guía intervenciones directas sobre el conductor, que se han venido desarrollando desde

hace años y que han mejorado la seguridad en las carreteras como, por ejemplo, la formación específica para el desarrollo de determinadas habilidades al volante o la selección de los conductores, rechazando a aquellos que no cumplan con un nivel mínimo de habilidad (Carbonell, et al., 1995).

En segundo lugar, **contribuyó al desarrollo** de la conocida como **Teoría de los Sistemas**, que toma como punto de partida la idea de que los accidentes se producen por algún tipo de desajuste en la interacción de los elementos del sistema complejo de tráfico. Esto trajo consigo un auge muy significativo, durante las décadas de los sesenta y setenta, de la modificación de los componentes técnicos del sistema de transporte por carretera (Elvik y Vaa, 2006).

Desde el punto de vista de esta teoría, si los humanos cometen errores en la conducción, estos se deben a un sistema que no estaba diseñado adecuadamente a sus capacidades. El trabajo de la Ingeniería de Carreteras y Vehículos en un esfuerzo máximo de adaptación de los componentes físicos del sistema a las capacidades humanas ha sido, sin duda, uno de los elementos que más han contribuido a la reducción de la tasa de accidentes por kilómetro recorrido en los países occidentales (Elvik y Vaa, 2006).

Aunque la mayoría de las medidas que se han tomado en los últimos años, para la mejora de la seguridad, derivan de esta visión de un modelo de habilidades, en la actualidad las líneas de investigación se orientan más a desarrollar un aspecto olvidado por este modelo y que es clave para entender la conducción como actividad compleja. Este aspecto tiene que ver con la conducta y el factor humano.

2.1.2 Modelos comportamentales

La visión de los modelos de habilidades dejaba en un segundo plano a los factores de tipo motivacional y emocional. No se estaba teniendo en cuenta la libre intervención del conductor, ni su participación clave en el proceso de toma de decisiones.

Sin embargo, la investigación venía señalando la mayor implicación en los accidentes de los errores de reconocimiento y toma de decisiones frente a los errores de ejecución. Además, cada vez era más aceptado que las causas indirectas de los accidentes (alcohol, drogas, estrés, exceso de velocidad, etc.) eran causas

típicamente humanas y tenían una influencia clave sobre la accidentalidad (Montoro, et al., 2000).

Por lo tanto, parecía evidente que se debía poner el acento en el estudio de variables cognitivo-motivacionales de tipo comportamental que explicaran el porqué de las decisiones que toman los conductores, que a su vez tienen influencia directa en la ocurrencia o evitación de un accidente. No se puede olvidar que factores como la velocidad, la distancia de seguimiento, o los adelantamientos con poca visibilidad son función directa de una decisión del conductor. Independientemente de sus habilidades, es la persona que conduce quien, en última instancia, decide qué actuación ejecuta.

Esta nueva perspectiva favoreció el desarrollo de modelos y teorías explicativas más complejas.

Los primeros en introducir un modelo de este tipo, basado en la conducta del conductor, fueron los finlandeses Risto Näätänen y Heikki Summala, en una época en la que aún dominaban los modelos de habilidades (Carbonell, et al., 1995).

Su propuesta se definió como **Modelo de Riesgo Cero** (Näätänen y Summala, 1974). Estos autores proponen un nuevo modelo centrado en factores motivacionales y que, desde su punto de vista, marca un cambio relevante en el estudio del factor humano. En sus propias palabras, este modelo es relevante para la Seguridad Vial al poner su énfasis sobre aquello que el conductor hace en una situación determinada y no tanto en sus habilidades como conductor o en las condiciones del tráfico.

Su modelo se apoya en la competencia que existe entre los motivos excitatorios de la conducta, los cuales impulsan a un conductor a realizar determinadas acciones, y el riesgo subjetivo asociado a esa situación, que actúa como elemento inhibitorio de esa conducta.

Por tanto, el control de las acciones y de la conducta regresa al conductor y la conducción deja de ser una tarea de *pasos impuestos* para tratarse aquí como una tarea de *pasos autoimpuestos* en la que el conductor hace algo más que simplemente responder a estímulos.

Es el equilibrio entre los diferentes motivos excitatorios y el principal inhibidor, que es el *Control Subjetivo del Riesgo*, lo que permite una toma de

decisiones encaminada a una conducción segura. Cualquier desequilibrio de este proceso se traducirá en una toma de decisiones arriesgadas, favorecedoras de una conducción menos segura.

Ante motivaciones, más allá de la propia llegada al destino (motivo último del viaje), como por ejemplo, el *modelado comportamental*; la *autoafirmación personal*; el *placer de la conducción* o la *búsqueda consciente de emociones*, que ejercen como **motivos excitatorios**, el *Control Subjetivo del Riesgo* actuará como una respuesta anticipatoria de una amenaza o de un riesgo, provocando diferentes grados de miedo. De esta forma, este mecanismo actúa como **control inhibitorio** frenando la toma de decisiones arriesgadas.

En consecuencia, las medidas que propugna este modelo van encaminadas a la reducción del diferencial producido entre el riesgo subjetivo, percibido individualmente por cada conductor, y el riesgo objetivo real que existe para una determinada situación del tráfico. Una mayor información, con campañas de largo recorrido temporal, y la prevención, introduciendo mayor número de limitaciones y señales de peligro, contribuirán a la reducción del umbral de riesgo subjetivo. A su vez, la introducción de mejoras asociadas a la seguridad como pueden ser: mejora de las carreteras, cinturones de seguridad o carriles adicionales, ayudarán a disminuir el riesgo objetivo de una determinada situación del tráfico.

Unos años después, en 1982, el canadiense Gerald Wilde propuso la que ha sido la teoría, sobre la influencia del comportamiento humano en los accidentes, más relevante y que más controversia ha generado en los últimos años (Evans, 1985, 1991; Graham, 1982; Haight, 1986; Hoyes y Glendon, 1993; OCDE, 1990).

Esta teoría, denominada **Teoría de la Homeostasis del Riesgo,** se incluye en los modelos comportamentales porque, aunque identifica la conducción como una actividad que implica la puesta en marcha de determinadas habilidades, es la motivación la que dirige el proceso (WIlde, 1982, 1988).

El modelo parte de dos premisas principales:

1. Las personas fijan un nivel de riesgo asumido, de acuerdo con las ventajas e inconvenientes percibidos respecto a la ejecución de conductas seguras.

2. La frecuencia de muertes y lesiones depende del estilo de vida y se mantienen en el tiempo, siguiendo un proceso de control autorregulador de ciclo cerrado. El planteamiento es que cada sociedad tiene la tasa de accidentes que desea, esto es, la tasa que la propia sociedad está dispuesta a asumir.

El propio autor explica la propuesta con la siguiente frase: *Dadme una escalera el doble de estable y subiré el doble de alto*. Mientras el nivel de riesgo aceptado permanezca estable, cualquier medida que introduzca una mejora en la seguridad originará algún tipo de ajuste comportamental que compensará dicha mejora para mantener el nivel de riesgo deseado.

De acuerdo con esta visión, es por esto por lo que algunas medidas, principalmente técnicas, no tienen los efectos deseados sobre la reducción de la tasa de accidentes. Por ejemplo, en Alemania el número de accidentes sufridos por taxis con sistema antibloqueo de frenos (ABS), en la época en la que se empezó a implantar este mecanismo, no fue menor que el número de accidentes de aquellos que no estaban provistos de este sistema. Se pudo comprobar que los conductores de vehículos con ABS redujeron las precauciones a la hora de conducir. Parecía como si estos taxistas pensaran [...] *para qué voy a tomar precauciones si mi coche frena mucho mejor y puedo evitar el accidente* [...]. Su conducta compensaba la mejora técnica manteniendo, de esta forma, estable el nivel de riesgo aceptado (OCDE, 1990).

En cierto modo, cada sociedad tiene el número de accidentes que está dispuesta a asumir. Por tanto, las medidas que este modelo propugna como efectivas son aquellas que van dirigidas a la reducción del nivel de riesgo aceptado por la sociedad.

No es objeto de este texto desarrollar todo el debate posterior al planteamiento de este modelo; baste aquí mencionar que la principal crítica que se le plantea es la imposibilidad de refutar la teoría (Elvik y Vaa, 2006). En el momento en que una medida concreta no es efectiva y no se reduce el número de accidentes, la teoría se confirma ya que mostraría que, como se ha explicado anteriormente, a pesar de la medida tomada, la conducta se ha ajustado para mantener constante el nivel de riesgo aceptado. Por el contrario, si una medida concreta acredita ser efectiva también la confirma, ya que se puede argumentar que la reducción de los accidentes se ha producido por una reducción en el nivel de riesgo aceptado. Así las

cosas, esta teoría sirve para explicar cualquier resultado. Sin embargo, es uno de los modelos que más ha influido en los últimos años de investigación en Seguridad Vial.

El último modelo al que se hará mención en este recorrido es el propuesto por Fuller en 1984. Es conocido como **Modelo de Amenaza - Evitación** (Fuller, 1984).

Partiendo de una revisión de los aspectos motivacionales del Modelo de Riesgo Cero de Näätänen y Summala y de la Teoría de Homeostasis del Riesgo de Wilde, el autor desarrolla un análisis de la conducta en términos de un modelo de amenaza-evitación que trata de explicar el comportamiento del conductor. El planteamiento pasa por analizar que, cuando el sujeto se enfrenta con un estímulo discriminante de un suceso potencialmente peligroso, la actuación del conductor se basa en la historia de recompensas y castigos que ha aprendido previamente.

Por lo tanto, en este modelo, los procesos de condicionamiento clásico y del aprendizaje instrumental son fundamentales (Carbonell, et al., 1995). A partir de ellos se puede explicar cómo el conductor conjuga las dos motivaciones principales que guían su actuación y la toma de decisiones:

1. La motivación por llegar cuanto antes a su destino.

2. La motivación por llegar sin haber sufrido daño, esto es, habiendo evitado los posibles riesgos.

Para el autor es el esquema de conducta ABC (Antecedents-Behavior-Consequences) el que mejor representa el proceso de toma de decisiones de un conductor, tras el aprendizaje de las correspondientes respuestas de evitación, bien sean *anticipatorias*, dirigidas a evitar la aparición del riesgo, o *retardadas*, en las que el margen de tiempo para reaccionar es menor, por lo que están más relacionadas con estilos de conducción más arriesgados.

De esta manera, el aprendizaje de la conducción se fundamenta en un proceso de refuerzo negativo, donde el sujeto aprende a asociar peligros con estímulos discriminantes y a ejecutar conductas preventivas que permitan evitar la ocurrencia del riesgo.

La intervención ante los accidentes desde este modelo pasa, entonces, por una mejora en las carreteras en términos de mejora de los estímulos discriminantes (señalización de tráfico), para facilitar las respuestas anticipatorias de evitación del

riesgo. Por otro lado, el entrenamiento en el manejo del vehículo en determinadas situaciones de riesgo mejoraría la eficacia de las respuesta. Este entrenamiento debe ir acompañado de un sistema claro de recompensas y castigos que favorezca las respuestas de anticipación frente a las demoradas.

Cualquiera de los modelos aquí descritos y aquellos que no se incluyen en este texto, como por ejemplo el modelo jerárquico de Bötticher y Van der Molen (van Der Molen y Bötticher, 1988), pueden servir de base para el establecimiento de algún proceso de selección de conductores.

Sin embargo, aún a pesar de todos estos avances, la poca evidencia que se ha encontrado que relacione de manera causal alguna variable clásica de personalidad con accidentalidad y el auge, hasta bien entrados los años setenta de los planteamientos de los modelos de habilidades, han hecho que la mayor parte del trabajo realizado en la selección psicológica de conductores se haya centrado en estudios de tipo psicotécnico, que evalúen variables como tiempo de reacción, atención y resistencia a la monotonía o percepción de velocidad y distancias.

En el uso y aplicación de esta metodología, España ha sido país pionero, hasta el punto de que puede hablarse de un *Modelo Español* de selección de conductores (Montoro, Carbonell, y Tortosa, 1991; Tortosa, Esteban, Tejero, y Molero, 1995; Tortosa y Montoro, 2002).

2.2 El Modelo Español de selección de conductores

Si en el momento actual se puede hablar de una tradición española en la aplicación de la Psicología a la Seguridad Vial y a la conducción, sus orígenes se encuentran, sin ningún género de dudas, en los trabajos de Emilio Mira, desde el Instituto de Orientación y Selección Profesional de Barcelona, y de José Germain, desde el instituto homólogo en Madrid (Tortosa, et al., 1995).

Contemporáneamente existían estudios e investigadores a nivel mundial que empezaban también a trabajar en estos aspectos. No debe olvidarse que los primeros trabajos sobre Psicología Experimental de Wundt coincidieron en el tiempo con el nacimiento del automóvil, pero también Patrizzi y Gemelli en Italia; Loemental, Moede o Piorkowski, entre otros, en Alemania; Münsterberg y Gerhardt, en Estados Unidos; Claparede, en Suiza o Manning, en Gran Bretaña , se interesaron por desarrollar una disciplina psicotécnica (Montoro, et al., 1991). Es evidente que la problemática que tiene implicada actualmente a la Psicología de la Seguridad Vial en la búsqueda de pruebas y evaluaciones válidas y predictivas ya se había presentado a principios del siglo XX (Germain, Pinillos, y Pascual, 1959).

Pero el desarrollo de esta disciplina en España, desde la celebración en Barcelona de la **I Reunión de la Asociación Internacional de Psicotecnia**, ha sido muy significativo. No en vano, en España, el rol de psicólogo de Seguridad Vial, como facultativo que examina la aptitud de un candidato a conductor, se creó antes incluso de que existiera la Psicología como una disciplina de pleno derecho y se ha mantenido desde entonces (Tortosa y Montoro, 2002).

El hilo conductor de todo este desarrollo ha sido el examen psicotécnico denominado *Selección psicológica de conductores* (Tortosa, et al., 1995) que, con muy ligeras modificaciones, aún hoy permanece vigente (Montoro, et al., 1991) desde que en 1922 el Ayuntamiento de Barcelona incluyera en la concesión de la explotación del servicio urbano de autobuses la obligatoriedad de que sus conductores obtuvieran un certificado de su aptitud profesional.

Fue el propio Mira quien estableció esas primeras pruebas de aptitud partiendo de su definición de lo que era la tarea de la conducción:

> Trabajo psicofísico espacial, variable, de grandes dimensiones, que requiere: *excelente apreciación de distancias y velocidades*; buena sensibilidad articular y muscular; buena percepción del tiempo; atención visual y auditiva extensa, intensa y persistente; *excelente control emocional*; rapidez y precisión de las reacciones visomotrices de manos y pies (Mira, 1928).

Así, esta selección de conductores profesionales se centraba en aspectos relacionados con la *percepción* y en la *asociación de estímulos percibidos* más que en la reacción, analizando la capacidad del candidato para percibir velocidades y distancias; la distribución, intensidad y duración de la atención visual y la rapidez y precisión de las reacciones oculomotrices y audiomotrices (Tortosa y Montoro, 2002).

Desde el Instituto de Orientación Profesional, este tipo de selección se extendió a Madrid y empezó a ser aplicada también por el equipo de Germain. El éxito de estos procesos selectivos generó un interés social por extender su aplicación a todos los conductores y no utilizar estas innovaciones únicamente con conductores profesionales. Esta pujanza social se tradujo en la inclusión en el Reglamento de Circulación que entró en vigor en 1934, de una evaluación psicológica obligatoria para todos los conductores de carácter periódico y sometida a una concesión limitada de los permisos. Esta iniciativa puso a España a la vanguardia de la implantación de medidas de carácter científico que contribuyeran a la mejora de la seguridad de los conductores (Tortosa, et al., 1995).

Durante el periodo posterior, a causa de la Guerra Civil Española y el durísimo periodo de postguerra, el desarrollo del modelo se estanca. Por un lado, no existen, con el nuevo Régimen, modificaciones en su normativa reguladora ni cambios en la

aplicación de las pruebas. Por otro, la actividad investigadora y la producción científica se detienen (Tortosa y Montoro, 2002).

No será hasta los años cincuenta cuando, en virtud de un significativo aumento del parque móvil, así como de los nuevos aires de apertura y modernización, la Psicología comience su institucionalización. De esta forma, la actividad de los Institutos de Orientación, que habían subsistido durante todo el periodo de postguerra gracias al trabajo de selección, se benefician de un proceso de amplia producción científica que va provocando un giro, desde una Psicotecnia objetiva, ligada al modelo de predisposición al accidente, hacia una evaluación de aptitudes más globales, como son la personalidad o la inteligencia (Tortosa y Montoro, 2002).

Todo este proceso de producción científica, validando y adaptando pruebas ya existentes en otros países, tiene como resultado la creación, a principios de los años setenta, de la batería del Instituto Nacional de Psicología Aplicada y Psicotecnia (INPAP), que incluía la evaluación de tres aspectos principales *Inteligencia, Aptitudes y Personalidad* (Germain, Pinillos, García, y Aberasturi, 1970).

A estas alturas, en el reconocimiento previo a la expedición de un certificado de aptitud para la conducción, las pruebas psicotécnicas evaluaban los siguientes aspectos (Tortosa, et al., 1995):

1. Atención distribuida y concentrada.
2. Precisión en la concepción de las diferencias de velocidad.
3. Coordinación de los movimientos de ambos brazos.
4. Rapidez, precisión y regularidad del tiempo de reacción simple y con inhibición.

Aún seguía considerándose la evaluación desde el punto de vista de la predisposición al accidente, tomando la conducción como una tarea reactiva, pero, el auge en los setenta de la Psicología como disciplina y el importante aumento de la siniestralidad en carretera llevaron a tomar, cada vez más en cuenta ,el papel del psicólogo como elemento clave en la mejora de las condiciones y seguridad del tráfico.

A principios de los años ochenta los Institutos de Psicología aplicada pierden su competencia sobre los exámenes selectivos de conductores. Con el **Real Decreto 1467/1982** se autoriza la creación de centros de reconocimiento privados que, desde entonces, son los que vienen realizando las evaluaciones psicológicas de los conductores profesionales.

Posteriormente, en 1985, el Real Decreto 2272 sustituye al anterior, extendiendo la exploración médico-psicológica a todos los permisos (Tortosa, et al., 1995). Para la obtención de un permiso de autobús (Clase D) este Real Decreto señala como obligatorias las siguientes pruebas:

- Anticipación de la velocidad.
- Coordinación bimanual visomotriz a ritmo impuesto.
- Tiempo de reacción múltiple discriminativo.
- Atención concentrada y vigilante a la monotonía.

Con carácter opcional también puede evaluarse:

- Capacidad de aprendizaje psicomotor.
- Personalidad.
- Estilos cognitivos.

Los centros de reconocimiento estaban siendo regulados, en parte, por el Real Decreto 1467/1982, de 28 de mayo, que determina las enfermedades y deficiencias que pueden impedir la obtención del permiso, restricciones y condiciones de expedición de los certificados de aptitud y reconocimientos correspondientes, en parte por el Real Decreto 2272/1985, de 4 de diciembre, modificado por el Real Decreto 1342/1986, de 26 de mayo, que regula las aptitudes psicofísicas que deben poseer los conductores de vehículos y que también regula las condiciones del ejercicio profesional de los centros de reconocimiento destinados a verificarlas.

La antigüedad de estos decretos (todos ellos son anteriores al texto articulado de la *Ley sobre Tráfico, Circulación de Vehículos y Seguridad Vial* (RDL 339/1990)) y la dispersión normativa en este ámbito han generado cierta inseguridad jurídica.

Por otro lado, el Real Decreto 818/2009, de 8 de mayo, que regula el nuevo Reglamento General de Conductores, recoge, en su anexo IV, las aptitudes psicofísicas requeridas para obtener o prorrogar la vigencia de los diferentes permisos de conducción conforme a la Directiva 2006/126/CE.

Todo este contexto normativo, unido a la relación de este tipo de centros con el Sistema Nacional de Salud, ha impulsado la redacción de un reglamento, aprobado por el Real Decreto 170/2010, de 19 de febrero, por el que se regulan, en una única disposición normativa, todos los aspectos relacionados con los centros de reconocimiento. Este Real Decreto deroga a los anteriores relacionados con este ámbito.

Las principales novedades, incluidas en este nuevo articulado, no afectan a las pruebas psicotécnicas que se aplican; van más en la línea de implantar mejoras administrativas, facilitando la gestión de los centros, en cuanto a autorizaciones y requisitos, y mejorando los sistemas de comunicación para hacer más sencillo el trámite al usuario. La mejora más importante es la posibilidad de la gestión telemática de los informes, lo que permite que sea el propio centro el que tramite, directamente con la administración, la renovación o expedición del permiso, sin que el usuario tenga que presentarse físicamente en las oficinas de la Jefatura de Tráfico.

La derogación de la normativa vigente hasta mayo de 2010 deja recogidos todos los aspectos relacionados con las aptitudes psicofísicas de los conductores en el anexo IV del Real Decreto 818/2009, al que se hizo referencia anteriormente. En él se estipulan los requisitos y límites de las diferentes aptitudes que se consideran imprescindibles para la conducción. En concreto, en el área perceptivo-motora se señalan cuatro aspectos:

1. **Estimación del movimiento:** No se admite ninguna alteración que limite la capacidad para adecuarse, con seguridad, a un sistema de tráfico que requiera estimaciones de relaciones espacio-temporales. A partir del Real Decreto 1467/1982 esta dimensión se evalúa con el *Test de anticipación de velocidad* de Maruyama y Kitamura (1961).

2. **Coordinación visomotora:** Alteraciones que supongan la incapacidad para adaptarse adecuadamente al mantenimiento de trayectorias establecidas. El procedimiento utilizado es el de *Doble*

Laberinto de Bonnardel (1946), con la novedad de la introducción del ritmo impuesto.

3. **Tiempo de reacciones múltiples:** No se admiten alteraciones graves en la capacidad de discriminación o en los tiempos de respuesta. En la última época se viene midiendo esta dimensión mediante *Polireactígrafos*, que evalúan tiempo de reacción y frecuencia de errores de respuestas motrices múltiples.

4. **Inteligencia práctica:** No se admiten casos en los que la capacidad de organización espacial resulta inadecuada para la conducción.

En lo referente a la **Personalidad**, la norma recoge que no deben existir trastornos graves, en particular, aquellos que se manifiesten en conductas antisociales con riesgo para la seguridad de las personas.

Por último, además de lo indicado para la inteligencia práctica, el Reglamento General de Conductores regula también los límites referentes al desarrollo intelectual, no debiendo existir un retraso mental con cociente intelectual inferior a 70.

Todo este recorrido de investigación y desarrollo normativo ha conducido al sistema actual, por el cual cualquier conductor que quiera obtener una autorización administrativa que le habilite para conducir un vehículo autocar debe cumplir tres requisitos fundamentales:

- Tener 21 años cumplidos para obtener un permiso D1 ó 24 años cumplidos para obtener un permiso D.

- Superar las pruebas prácticas y exámenes teóricos realizados por la Dirección General de Tráfico dirigidos a evaluar unos criterios mínimos.

- Superar la evaluación de aptitudes y capacidades básicas para la conducción recogidas en el modelo español de reconocimiento psicotécnico.

A pesar de que, administrativamente, este proceso es la vía de acceso reconocida a una autorización legal para manejar un vehículo, en forma de permiso de conducción, en el caso de los carnés de autobús y autocar, desde el punto de

vista empresarial, el sistema se queda corto a la hora de seleccionar a aquellos conductores con mejores aptitudes y habilidades para ejercer la profesión con unas garantías mínimas de seguridad y confort para el viajero.

Si una empresa de transporte público de viajeros requiere de los servicios de un conductor, no puede fiar su contratación exclusivamente al simple cumplimiento de unos requisitos administrativos o a la posesión de un permiso. Necesariamente hará pasar al candidato por una prueba práctica de conducción real, con un vehículo similar al que manejará durante su trabajo.

Una evaluación práctica de este tipo es, sin duda, un método aceptado para determinar las capacidades y estilo de conducción de un conductor (Justiss, 2005). El principal problema radica en cómo estandarizar ese proceso de forma que proporcione información clara, interpretable y objetiva.

2.3 ¿Por qué es necesaria una evaluación práctica de los conductores profesionales?

El uso más habitual y extendido entre las empresas de transporte de viajeros, para el proceso de toma de decisiones a la hora de incorporar a un conductor a su plantilla, es el de realizar al candidato una evaluación práctica de sus habilidades al volante. En general, durante la prueba el aspirante se enfrenta a un recorrido definido previamente y con diferentes niveles de dificultad a los mandos de un vehículo similar al que utilizaría en su trabajo diario, en el caso de ser seleccionado, mientras es evaluado por un *experto* de la empresa (generalmente el jefe de taller o un jefe de tráfico con experiencia anterior en conducción).

Tal es la importancia de esta práctica que, en la mayoría de las ocasiones, lo más común es que sea el único elemento de decisión del que dispondrá la empresa para decidir la contratación del candidato. Sin embargo, no existe una prueba única, estandarizada y objetiva que, conocidas sus propiedades psicométricas, pueda asegurar una evaluación correcta de los aspirantes y favorecer el éxito del proceso de selección.

Este tipo de evaluación *al volante* ha sido definida por diversos autores como el método más adecuado para determinar la competencia y las habilidades de conducción de una persona (Hunt, et al., 1997; Odenheimer, et al., 1994). Aún así, este sistema no está suficientemente estandarizado. No se ha encontrado una única prueba de aplicación general que pueda ser utilizada por cualquier empresa o investigador interesado. En las pruebas existentes, los métodos de puntuación y

procedimientos no son comunes y varían con mucha frecuencia; tampoco hay demasiados estudios de sus propiedades psicométricas. En su aplicación al mundo empresarial, esto lleva a que cada organización tenga sus propios métodos sin que hayan pasado por un proceso de estudio mínimamente profundo, en cuanto a instrumento de medida. Al final, lo habitual es que, exista prueba de evaluación o no, sea el criterio del *experto* elegido por la empresa el que decida sobre la aptitud del candidato. El Apto o No Apto definitivos no se otorgan por medio de una evaluación objetiva y estandarizada.

En la literatura, el desarrollo de herramientas de este tipo se ha centrado principalmente para la evaluación de personas de edad avanzada o con algún tipo de impedimento, ya sea por una enfermedad degenerativa o por un evento traumático. Han sido la Terapia Ocupacional y la Geriatría, desde la importancia que se le otorga a la capacidad de conducir como actividad clave para la vida diaria y su relación con una posible pérdida de libertad e independencia, quienes han trabajado más en profundidad sobre el desarrollo de este tipo de instrumentos de evaluación de las habilidades al volante para, de esta forma, poder determinar si una persona reúne las condiciones necesarias para conducir de manera habitual (Justiss, 2005).

Todas las pruebas desarrolladas en este ámbito de trabajo con ancianos o personas impedidas se centran en conductas al volante relacionadas con tareas específicas, bajo unas determinadas condiciones ambientales (Justiss, 2005).

El más utilizado es el Washington University Road Test (WURT), desarrollado en 1997 para evaluar conductores con demencia (Hunt, et al., 1997) y que, posteriormente, ha sido utilizado por diferentes autores en diversos estudios (Carr, Barco, Wallendorf, Snellgrove, y Ott, 2011; Justiss, 2005 ; Roe, et al., 2003; Vrkljan, McGrath, y Letts, 2011).

En esta prueba, un instructor autorizado guía a los sujetos por un recorrido fijo, previamente establecido. Se puntúan 54 conductas en una escala de tres niveles. A esto se añade una puntuación final subjetiva en la que el evaluador manifiesta su impresión global de la ejecución.

Otras pruebas de evaluación han sido desarrolladas expresamente para la realización de estudios concretos. Tal es el caso de la prueba utilizada para el estudio de Richardson y Marottoli sobre la relación entre la atención visual y las conductas

en conducción (Richardson y Marottoli, 2003). En este caso, los autores evalúan 36 conductas sobre la base del Apto/No Apto. Toda vez que se construyó específicamente para el desarrollo de una investigación concreta, esta prueba no aporta un resultado global que, al final, pueda ser comparado con los resultados de una evaluación sobre las habilidades de conducción.

Anteriormente otros autores también habían desarrollado sus propios instrumentos de evaluación *al volante*. El **Performance-Based Driving Evaluation** (PBDE) de Odenheimer (Odenheimer, et al., 1994) utiliza una ruta fija y el evaluador asigna puntos en una escala de 4 niveles a 68 tareas relacionadas con la conducción, las cuales habían sido previamente determinadas por un grupo de expertos, como las más apropiadas para evaluar las habilidades propias de una conducción segura.

En general, estas pruebas presentan unos buenos resultados, en cuanto a consistencia interna y fiabilidad interjueces. En cuanto a su validez, la mayoría de ellas se asocian a tareas concretas de memoria, tareas visuales y puntuaciones globales en test de funciones mentales, con unos buenos resultados. Todos están correctamente diseñados y logran evaluar aspectos relevantes para la problemática de la ejecución de tareas durante la conducción. Sin embargo, no establecen un protocolo único de evaluación que pueda proporcionar información suficiente para clasificar correctamente las actuaciones seguras frente a las no seguras. Además, están muy centradas en la evaluación de personas con edad avanzada o discapacidades cognitivas.

La experiencia con este tipo de pruebas demuestra que es necesario establecer una evaluación estandarizada que pueda ser universalmente aceptada.

Existen herramientas que buscan esa universalización dirigida al conductor en general, con independencia de su edad o su capacidad para la conducción o el procesamiento de la información. Sin embargo, desde esta perspectiva se ha trabajado poco con un enfoque científico. La mayoría de estas herramientas de evaluación tienen una clara orientación comercial, lo que, en mi opinión, las convierte en instrumentos difícilmente accesibles para la comunidad científica y sobre los que los estudios de resultados son, cuanto menos, opacos.

En Canadá está muy extendido el uso de una herramienta conocida como **DriveAble**. Combina una evaluación de competencias, realizada por ordenador y que mide atención, memoria, reflejos y toma de decisiones, con una evaluación

práctica que registra 34 maniobras de dificultad creciente (Dobbs, Heller, y MacDonald, 1998).

También es conocido un producto llamado **Virtual Risk Manager**, comercializado por Interactive Driving Systems. Como herramienta de evaluación, se asegura que puede detectar aquellos conductores con potencial riesgo de padecer un accidente (Murray y Cuerden, 2006).

Herramientas comerciales de este tipo presentan el problema de que, por el afán proteccionista de un interés comercial, resulta complejo para la comunidad científica poner a prueba sus propiedades, su utilidad real como instrumento de medida y su validez para poder ser utilizada como criterio sobre el que tomar decisiones posteriores.

Normalmente, las informaciones que se dan sobre este tipo de herramientas se apoyan en una estrategia comercial caracterizada de una parte, por interpretaciones muy positivas de los datos estadísticos obtenidos y, de otra, en una aparente autoridad científica. Sobre este particular puede consultarse un interesante artículo del profesor Anders af-Wåhlberg de la universidad de Uppsala (af-Wåhlberg, 2007).

Poco o nada se ha hecho desde el punto de vista de la selección de personal para desarrollar una herramienta de este tipo que permita determinar el nivel de habilidad de un conductor profesional. Por extensión, no existe ninguna prueba de evaluación, específica para conductores de autobús, que haya sido estudiada desde el enfoque de sus propiedades psicométricas y que esté dotada de las mínimas características de estandarización para hacer más objetiva y generalizable su aplicación, de manera que pueda servir como criterio de selección de personal previo a su incorporación a un puesto de trabajo como conductor.

Ante esta situación, el principal operador de transporte de viajeros en España, con una plantilla de más de 6.000 trabajadores y por encima de las 300 incorporaciones anuales de nuevos conductores, se plantea que las pruebas de conducción que habitualmente venía realizando desde su departamento de selección debían pasar por un proceso de renovación que, necesariamente, incluyera un estudio en profundidad de su procedimiento de aplicación y puntuación, así como de sus propiedades como instrumento de medida para, de esta forma, poder estandarizar el proceso y hacerlo de aplicación común en otros

puntos de la empresa por evaluadores adiestrados, mejorando de esta forma el proceso global de selección. Fruto de este interés, surge este trabajo de tesis doctoral.

Tiene mejor conocimiento del mundo, no el que más ha vivido, sino el que más ha observado.

Arturo Graf

Construcción, Descripción y Condiciones Generales para la aplicación de la Prueba

3

En este capítulo se describe el proceso de construcción de la prueba, desde los documentos que existían en origen, hasta el panel de expertos por el cual se determinó la prueba definitiva.

Se explicitan las características generales del test y sus condiciones de aplicación, definiendo los diferentes aspectos que deben ser considerados, a la hora de evaluar cada uno de los ítems.

Por último, se plantean los objetivos generales que se persiguen con el desarrollo de la investigación y las hipótesis de trabajo para esta tesis doctoral.

3.1 Construcción y descripción de la Prueba Práctica de Evaluación de Conductores

Lo que aquí se recoge es el fruto del trabajo de varios años en los que el Departamento de Recursos Humanos de ALSA, desde su departamento de Selección y Formación, ha tratado de estandarizar y hacer más objetivo todo el proceso de selección de conductores y, para este caso en concreto, la evaluación de las aptitudes de conducción de los candidatos que acuden a pruebas de selección para ocupar puestos de conductor.

3.1.1 Ficha técnica

- **Nombre original**: Evaluación Práctica de Conducción.
- **Autores**: Servicio Horizontal de RRHH de ALSA, Dpto. de Selección y Formación.
- **Procedencia**: Elaboración propia de los autores.
- **Tipo de Administración**: Individual en entorno práctico de conducción.
- **Metodología**: Observacional.
- **Duración**: Entre 45 minutos y 1 hora, incluyendo instrucciones previas y comentarios finales.

- **Requisitos**: 24 años en adelante. Candidato en posesión de permiso clase D en vigor. Vehículo autocar de 12 metros.

- **Nivel cultural mínimo**: Graduado escolar o similar.

- **Puntuación**: Escala de 4 niveles. Máximo 39 puntos. Apto en 23 puntos

- **Significación**: Evalúa las habilidades y aptitudes del candidato como conductor de vehículo autocar en cuanto a: Conducción Económica, Seguridad y Conducción Técnica.

- **Tipificación**: El estudio Psicométrico se realizó sobre una muestra representativa de conductores españoles, se ha elaborado la baremación correspondiente.

3.1.2 Criterios de elaboración

La prueba de Evaluación Práctica de Conducción (en adelante EPC) fue elaborada inicialmente a partir de las pruebas para evaluación de conductores que se venían haciendo dentro del Departamento de Selección y Formación de ALSA, desde su puesta en marcha en 1986.

El objetivo era tener un instrumento estandarizado, común para todos los evaluadores, que recogiera los aspectos relevantes que un conductor profesional debe dominar, en mayor o menor medida, para ejecutar la práctica diaria de su trabajo con garantías suficientes de seguridad y comodidad para los viajeros.

Ese primer instrumento, creado de manera intuitiva, sin reparar excesivamente en aspectos psicométricos propios de la elaboración de una herramienta de medida, se convirtió en parte fundamental del proceso de selección, ámbito para el que fue concebido inicialmente.

Sin embargo, la revisión a la que fue sometido para la elaboración de la actual EPC, revela que puede ser utilizada eficazmente, no sólo para procesos de selección de candidatos externos a la Organización, sino también como herramienta de evaluación para seguimiento de la progresión en cursos de formación de larga

duración, así como en sesiones formativas relacionadas con la Seguridad y la Conducción Eficiente.

En el proceso de revisión y creación de la EPC se ha tratado de elaborar un cuestionario observacional coherente con los requisitos mínimos exigibles a un conductor profesional de autobús. Se ha procurado facilitar la interpretación de los diferentes ítems, así como la definición de los distintos niveles, para hacer más objetivo el proceso de cumplimentación de la hoja de evaluación, de acuerdo con lo observado.

La prueba está concebida de manera genérica y permite evaluar al candidato con independencia del tipo de trabajo, urbano, regional o largo recorrido que vaya a realizar. Eso sí, se debe modificar la puntuación de aptitud final del candidato para servicios urbanos, ya que existen, al menos, tres criterios que la prueba evalúa y que no son relevantes para la ejecución de un conductor en servicios urbanos de transporte.

A la hora de establecer los diferentes ítems y elegir el contenido de la prueba de evaluación se tomó como punto de partida una prueba que ya existía, así como los criterios que se entendían como más importantes, en relación con la tarea que se va a ejecutar, constatados en la observación directa del ejercicio profesional de la ocupación.

Se constituyó un panel de expertos con aquellos evaluadores del Departamento de Selección y Formación que venían utilizando habitualmente la herramienta antigua. De las reuniones de trabajo de ese panel de expertos surge el consenso de los contenidos de la prueba, basándose en tres aspectos principales: *Conducción Económica, Seguridad y Nivel técnico de conducción*.

La determinación de las aptitudes que se debían evaluar constituyó la parte más compleja del proceso de elaboración de la prueba. Los criterios que se establecieran deberían recoger todas las habilidades que un conductor pone en práctica en el desarrollo diario del trabajo y a su vez predecir, con eficacia, el éxito de los candidatos en su desempeño futuro como trabajadores.

En este proceso, el conocimiento de la profesión y la experiencia práctica, como evaluadores de los participantes en el panel de expertos, fueron puntos clave para determinar aquellos criterios que debían mantenerse y aquellos susceptibles de eliminarse de la prueba que se venía utilizando anteriormente, así como los

nuevos aspectos a incluir y su relevancia, dentro de cada uno de los tres criterios generales considerados.

Como punto de partida, se consensuó y aceptó la definición de la habilidad en la conducción como:

> Capacidad de actuar que se desarrolla gracias al aprendizaje, al ejercicio y a la experiencia y que, para el caso de la conducción de vehículos, implica la ejecución de una serie de técnicas y la puesta en marcha de un esquema de conducta relacionado con la Seguridad y Eficiencia Energética.

Con esa orientación en mente, el panel de expertos trató de sintetizar en la prueba todas las evidencias conductuales que pudieran acreditar habilidad en la conducción, haciendo hincapié en los aspectos reseñados en la definición consensuada.

La construcción y elaboración de la prueba se ha realizado ateniéndose a los siguientes criterios de orden práctico:

- La prueba ha de poder realizarse de manera individual en una situación práctica de conducción.
- Cada criterio debe poder apreciar la intensidad con la que el candidato posee la aptitud a la que se refiere. Se debe buscar, más el grado de profundidad con la que el candidato puede llevar a cabo la tarea requerida y no tanto la facilidad o rapidez con la que la realiza.
- Los resultados se ofrecen diferenciando las puntuaciones absolutas de cada uno de los criterios, pero la puntuación total es la que marcará la aptitud o no del candidato para el ejercicio profesional.
- Se elaborará una tipificación y baremación para evitar trabajar directamente con puntuaciones absolutas.
- La hoja de control para el evaluador ha de ser sencilla en su manejo, para poder utilizarla dentro del vehículo donde se realiza la prueba, y ha permitir recoger toda la información necesaria para una cumplimentación rápida y eficaz

- Las instrucciones de aplicación y evaluación han de ser sencillas y homogéneas, especificando claramente las rutinas y tiempos a utilizar durante la prueba, así como definiendo las evidencias conductuales a observar y tener en cuenta para cada uno de los ítems del test.

3.1.3 Significación de cada ítem

Tras los paneles de expertos, la prueba definitiva quedó conformada de la siguiente manera:

Criterio 1: Conducción Económica

1.1 Control de las revoluciones por minuto del motor

1.2 Uso del retarder

1.3 Uso del freno de servicio

1.4 Anticipación

Criterio 2: Seguridad

2.1 Control emocional

2.2 Observación de los retrovisores

2.3 Distancia de seguridad

2.4 Manejo del volante

2.5 Atención a la señalización vial

Criterio 3: Conducción Técnica

3.1 Fluidez en la circulación

3.2 Observación y utilización del cuadro de mandos

3.3 Manejo del cambio

3.4 Sentido de la maniobrabilidad

A continuación se describen individualmente cada uno de los trece ítems que conforman la prueba.

3.1.3.1 Control de las revoluciones

Implica la medida del control de las revoluciones y la gestión eficiente de la potencia del motor.

La observación debe centrarse en comprobar que la aguja del cuentarrevoluciones está situada en la zona marcada en verde por el fabricante. Eso será indicativo de que se lleva un control eficiente de las revoluciones del motor y que se hace un aprovechamiento óptimo de la potencia del vehículo.

En una puntuación baja de la evaluación se sitúan las actuaciones en las que el conductor estira la marcha del cuentarrevoluciones continuamente, por encima de la zona óptima, llegando prácticamente hasta el corte de inyección; también se valorará de forma negativa la acción contraria, cuando cambie *bajo de vueltas*, sin llegar a la zona óptima. En el polo positivo se sitúa la ejecución de un candidato que mantiene sistemáticamente el régimen de marcha en la zona verde, optimizando los cambios y adaptándolos a las necesidades que marca la ruta.

La valoración de este ítem **no** formará parte de la puntuación total, cuando las pruebas se realicen en un vehículo autocar, tanto urbano como interurbano, **con cambio automático**.

3.1.3.2 Uso del retarder

Implica un uso racional y adecuado del mecanismo de retardo hidráulico de frenada.

Alcanzar un nivel óptimo de uso de este mecanismo implica hacerlo con antelación y de manera progresiva, accionando la palanca punto a punto, según necesidad, y haciendo uso, cuando sea posible, del programador. Adaptando así la velocidad del vehículo a las circunstancias del tráfico y de la vía. El uso correcto de este retardador reduce el número de veces que se utiliza el freno de servicio.

El desconocimiento del mecanismo supone una puntuación baja ya que se traducirá en la ausencia de uso. En el otro lado, cuando la utilización se ajuste a la forma arriba indicada y en los momentos adecuados, la puntuación será máxima.

La valoración de este ítem **no** formará parte de la puntuación total, cuando las pruebas se realicen en un vehículo autocar de **transporte urbano**, **con cambio automático**.

3.1.3.3 Uso del freno de servicio

Implica el uso adecuado del freno de servicio para detener el vehículo, en términos de suavidad de frenada y anticipación.

La observación de este ítem debe centrarse en el tipo de uso que el candidato hace del freno de servicio, atendiendo a criterios de suavidad y anticipación en la frenada. En el extremo más negativo se sitúan acciones en las que el freno es accionado repetidas veces, en situaciones en las que no es necesario y cuando se hace con mucha brusquedad. En la parte positiva, se valorará el que el uso sea suave, anticipativo y solo en los momentos en que es necesario, usando para ello una única frenada, evitando *clavar* el autobús en el momento de la detención total, premiando de esta forma el bienestar que se ofrece al viajero.

Es importante tener en cuenta que en la práctica habitual de uso de la herramienta, se han detectado varias diferencias de criterio con respecto a la evaluación de este punto. Una de las claves a la que se debe prestar especial atención es incidir en **no tener en cuenta la oportunidad de la frenada**, como elemento para valorar de forma negativa este ítem. En este punto, no se está valorando si es oportuno usar el freno de pie o si se pudiera sustituir la frenada de pie con el uso del retarder. Para evaluar correctamente esta parte del cuestionario, lo más importante es observar y valorar con detenimiento **el uso que se hace del freno de servicio y la calidad de la frenada resultante**. Realizar un número excesivo de frenadas con el freno de servicio ya está evaluado y penalizado en la evaluación *Uso del retarder*.

3.1.3.4 Anticipación

Implica valorar la capacidad del candidato de *mirar lejos*, mientras ejecuta la conducción. Como criterio general aceptado, se tomará que sea capaz de anticipar hechos del tráfico situados, como mínimo, a 15 segundos de su ubicación.

De esta manera, la capacidad de conducir con esa información permite al candidato detectar, sopesar y actuar en consecuencia sobre acontecimientos del

tráfico que están a 15 segundos *en su futuro*, pudiendo de esta forma anticipar con seguridad su actuación y la del resto de los usuarios de la vía. Esto redundará, no sólo en una conducción más segura, sino también en una mayor fluidez en la circulación y menos detenciones inesperadas ya que la práctica de esta aptitud mantiene la alerta sobre las circunstancias externas del tráfico y de la vía, así como la atención activa a los elementos relevantes, tanto fijos como móviles, de su entorno más inmediato.

La observación de su forma de circular, su capacidad para actuar con suficiente antelación ante las circunstancias y el tiempo de toma de decisiones serán puntos clave a tener en cuenta, a la hora de valorar este ítem.

3.1.3.5 Control emocional

Implica el dominio de uno mismo y de sus sentimientos. Está relacionado con la capacidad de hacer frente a una situación, potencialmente estresante, con aplomo y sin excesivo nerviosismo. Eso genera un clima de confianza y transmite sensación de seguridad.

Es habitual que los candidatos con escaso control emocional, por no tener experiencia previa con vehículo autocar o por no poseer estrategias de afrontamiento de situaciones estresantes, lo transmitan al puesto de conducción. Un conductor emocionalmente controlado se muestra resolutivo y decidido, no titubea al hablar, es dispuesto y no conduce transmitiendo nerviosismo, sino que lo hace de una manera pausada.

La evaluación de este ítem es la menos objetiva de toda la prueba, por lo que se deberá tratar de recoger el mayor número de evidencias conductuales que puedan sustentar el extremo de la decisión, tratando así de paliar la posible distorsión introducida.

3.1.3.6 Observación de los retrovisores

En este ítem se intenta evaluar, no sólo la frecuencia de observación de los retrovisores (lo recomendado es una vez cada 5 u 8 segundos) sino también el procesamiento de la información recibida.

El candidato deberá ser capaz de asimilar la información que le proporcionan los retrovisores, analizarla e integrarla para una toma de decisiones correcta. El uso e interpretación adecuada de esta información se traducirá en decisiones; aceleración, frenado, maniobras, señalización con intermitencias, incorporación a una nueva vía, etcétera, acertadas y en correcta relación con la información proporcionada por la visión de 180 grados hacia atrás que ofrecen los retrovisores, lo que permite tener una completa observación periférica.

No sólo se pide que el candidato tenga un uso y procesamiento correctos de la información de los retrovisores, sino que también deberá tener en cuenta el espacio físico que ocupan como prolongación de la carrocería del autobús.

3.1.3.7 Distancia de seguridad

Una distancia de seguimiento correcta, tanto al circular como al detenerse, contribuye de una manera muy significativa al incremento de los niveles de seguridad de un conductor.

La evaluación de este ítem implica comprobar que el candidato no se aproxima en exceso al vehículo precedente, al circular (en condiciones atmosféricas y de visibilidad óptimas, la distancia de seguimiento mínima debe ser de 1 segundo por cada 3 metros de vehículo, nunca siendo inferior a 5 segundos), asegurando así la posibilidad de realizar una frenada de emergencia en condiciones de seguridad.

Se debe prestar atención, también, a la distancia lateral que se mantiene circulando, tanto en adelantamientos como en la circulación urbana y tanto con respecto a vehículos circulando como a obstáculos fijos en la vía.

También, en las detenciones, se debe evaluar que la parada se hace a una distancia adecuada, con respecto al vehículo precedente (en condiciones normales, una distancia mínima equivalente al largo de un turismo se considerará adecuada) . De esta manera se protege de posibles golpes al vehículo precedente en caso de ser golpeado por detrás por otro vehículo, consiguiendo también espacio para maniobrar, en caso de avería del vehículo ante nosotros.

3.1.3.8 Manejo del volante

Implica comprobar que el candidato transmite seguridad en el uso del volante, haciéndolo girar con suavidad, pero a la vez sujetándolo con firmeza, no pasándoselo de una mano a otra ni dejándolo suelto aunque sea por intervalos cortos de tiempo, evitando malos usos, como puede ser el giro del volante en maniobra, utilizando únicamente la palma de una mano.

Se valorará que la postura de sujeción del volante y la posición de las manos con respecto al mismo es la adecuada, no sólo en línea recta, sino también cuando se circula en curva, especialmente si las curvas se ejecutan a alta velocidad.

Se debe prestar atención al trabajo y el manejo que el candidato hace del volante y cómo optimiza ese uso para aprovechar al máximo la ejecución de giros y maniobras. **Se trata de evaluar cómo interacciona con el instrumento que le permite dirigir el vehículo, no la calidad de la maniobra**.

Aunque la relación es directa, pueden darse casos en los que la maniobra se ejecute con éxito haciendo un mal uso del volante. En el extremo opuesto, un conductor con una perfecta sujeción del volante puede no ser capaz de ejecutar maniobras concretas. De nuevo, lo que se debe valorar es la corrección en la sujeción y la correcta postura con relación al volante; la corrección o incorrección de la maniobra se evaluará más adelante.

3.1.3.9 Atención a la señalización vial

La evaluación de este ítem implica valorar el cumplimiento y respeto del candidato hacia toda la señalización vial (marcas viales, señales verticales, regulación semafórica, agentes de la autoridad, etcétera) y, además, la comprobación de que es capaz de identificar, interpretar y orientarse, siguiendo la señalización informativa de población e indicadores de dirección.

El candidato debe respetar en todo momento la señalización vial, especialmente señales de *Stop* y *Ceda el Paso* en intersecciones, la regulación semafórica (infringir una de estas señales supondrá inmediatamente una puntuación de cero en este ítem) y los límites de velocidad, con mayor atención en zona urbana y periurbana; pero también, deberá ser capaz de orientarse y seguir las indicaciones leyendo la información pertinente en los carteles informativos, con

especial atención a los paneles indicadores de dirección y los informativos de rotonda.

3.1.3.10 Fluidez en la circulación

Una circulación fluida entre el tráfico implica, fundamentalmente, que el conductor sea capaz de ocupar, en cada momento, el carril de su conveniencia que ofrezca menor resistencia para su circulación.

De esta forma, se conseguirá una mayor visibilidad (lo que redundará en una mejora de la anticipación), mejor predicción de los posibles obstáculos de la vía y se evitarán las detenciones frecuentes.

La fluidez en la circulación se identifica en un transito confortable, sin interrupciones, combinando con habilidad el uso del volante, pedales, retarder y cambio de marchas, adaptada en todo momento a las condiciones cambiantes de la vía.

La puesta en práctica del sentido de la oportunidad para decidir realizar un adelantamiento, evitando quedarse retenido tras un vehículo más lento o estacionado perjudicando la circulación, o adaptar la marcha para no verse obligado a parar en un semáforo a punto de abrir, son otros indicadores importantes de la capacidad de realizar una circulación fluida.

3.1.3.11 Observación del cuadro de mandos

Implica el conocimiento y uso de los mandos e indicadores más habituales en cualquier vehículo autocar.

Es importante establecer la familiaridad de uso de los mandos **más habituales** (apertura de puertas, aviso de *warning*, intermitencias, luces interiores y exteriores, limpiaparabrisas, etcétera) chequeando su reconocimiento y agilidad de uso. Esto acredita experiencia en las operaciones habituales a bordo de un autobús y capacidad de adaptación a nuevos modelos.

Por otro lado, el conocimiento y uso de la información que proporcionan los indicadores del autocar es importante y acreditativo de capacidad y hábito de comprobar la información de retorno que proporciona el autocar, como por ejemplo las luces que se puedan encender en el panel de instrumentos.

Es importante, en este punto, valorar en su justa medida las diferencias que se puedan establecer entre los diferentes modelos de vehículo. No todos los botones y mecanismos están situados en los mismos lugares, pero los iconos identificativos son similares; por lo tanto, un candidato con experiencia debería ser capaz de adaptarse, en un tiempo razonable, y localizar los botones y mecanismos de acción más habituales.

3.1.3.12 Manejo del cambio

Implica llevar el vehículo en la relación de marcha apropiada en cada momento, efectuando los cambios de marcha con suavidad y sin ocasionar brusquedades al vehículo, lo que redunda principalmente en el comodidad del viajero.

El hecho de que suba o baje las marchas de una en una no es un punto crítico en la observación ni en la evaluación de la ejecución, siempre que la marcha sea la adecuada a las condiciones de la marcha y de la vía.

En este sentido, que el candidato realice el doble embrague **no será** puntuado positivamente.

La valoración de este ítem **no** formará parte de la puntuación total, cuando las pruebas se realicen en un vehículo autocar, tanto urbano como interurbano, **con cambio automático**.

3.1.3.13 Sentido de la maniobrabilidad

Implica capacidad de estimación de los espacios, las medidas del vehículo y los giros requeridos para ejecutar con precisión una maniobra.

El candidato debe poder estimar si la maniobra es posible, teniendo en cuenta las medidas del vehículo y el espacio disponible; debe poder trazar un plan de acción de cómo realizar la maniobra para conseguirlo y, por último, debe ser capaz de ponerlo en práctica, ejecutándolo con precisión a los mandos del vehículo.

Para evaluar correctamente este ítem es importante recabar evidencias que comprueben la capacidad del candidato de ver y desarrollar un plan de acción para acometer la maniobra, independientemente de su posterior ejecución técnica., puesto que establece claramente la diferencia entre aquellos que, siendo capaces de

trazar un plan de acción, técnicamente no son hábiles para ejecutarlo, de los que no son capaces de describir qué maniobra deberían poner en marcha.

Un último aspecto que aporta claves sobre la capacidad técnica de maniobra del candidato es el uso que haga de los voladizos, tanto delantero como trasero, aprovechando su paso sobre aceras y bordillos.

3.2 Condiciones generales de aplicación

En la creación y desarrollo de la hoja de evaluación de la EPC se ha tenido en cuenta la simplicidad y sobre todo la descripción clara y pormenorizada de los niveles que forman los diferentes ítems.

Al seguir una metodología observacional, la facilidad para trasladar la información obtenida a la plantilla de recogida de datos es un requisito muy importante que se ha tenido en cuenta durante todo el proceso de elaboración de la prueba. Se ha tratado, en la medida de lo posible, de definir con total claridad, no sólo los trece ítems que conforman el total de la evaluación, sino también de especificar claramente en la plantilla las conductas observables que establecen los límites de los cuatro niveles de evaluación. A tal fin, se ha intentando que estas descripciones no se solaparan, permitiendo así evaluar con total objetividad y sin dejar espacio para la duda a la hora de puntuar cada uno de los criterios.

3.2.1 Indicaciones sobre el desarrollo de la prueba

Para la aplicación de la prueba se tendrán en cuenta las normas propias de toda evaluación práctica:

- La prueba deberá ser efectuada, en la medida de lo posible, cuando el candidato no esté cansado y se le deberá advertir de que, durante el desarrollo de la evaluación, no se permitirá que sea requerido para ninguna otra actividad. La recomendación en este sentido es que

lleve su teléfono móvil apagado durante la realización de toda la prueba, para evitar interrupciones.

- Se procurará que el vehículo con el que se vaya a realizar la prueba sea lo más similar posible al vehículo con el que se realizará el trabajo posterior, teniendo en cuenta que **nunca** deberá realizarse la prueba en vehículos inferiores a 12 metros. Esto es especialmente importante cuando se trata de vehículos de 15 metros o aquellos específicos para el transporte urbano. Si esto no fuera posible, a la hora de cumplimentar la hoja de evaluación se tendrán en cuenta las limitaciones que se hayan dado con respecto al vehículo.

- Los itinerarios seleccionados para la realización de la prueba deberán estar adaptados a diferentes niveles de dificultad, conforme a la experiencia del candidato, y tener una ruta establecida que permita verificar, en situaciones de tráfico real, todos los criterios contenidos en la evaluación. Se procurará incluir un ejercicio de maniobra de aparcamiento.

- Obligatoriamente, antes del inicio de la evaluación, todos los candidatos habrán de identificarse debidamente y entregar al evaluador su permiso de conducción en regla y su tarjeta CAP, aquellos que la requieran, acreditando de esa manera su aptitud administrativa para realizar la prueba.

- Se les entregará, antes de ponerse al mando del vehículo, un disco diagrama de tacógrafo para su cumplimentación y uso. En vehículos con tacógrafo digital se les pedirá que inserten su tarjeta personal de tacógrafo. El conocimiento del uso del tacógrafo no es un aspecto que se valore en la evaluación, pero si se hará mención del hecho si el candidato manifiesta no saber utilizarlo o no lo utiliza o cumplimenta correctamente.

- El evaluador, que estará presente durante toda la prueba, se situará ocupando la plaza cuatro del autocar; en autobuses urbanos, se situará en una posición similar a la que tendría ocupando dicha plaza en un interurbano. Tendrá en su poder, discos-diagrama para el tacógrafo y hojas de evaluación suficientes. En todo momento esta

documentación debe permanecer en lugar no visible para el resto de los evaluados.

- Se solicitará que todas las aclaraciones sobre el desarrollo de la prueba o dispositivos del autobús se hagan antes de empezar. Responder de modo individual, tanto al candidato al volante como al resto de los asistentes, circulando con tráfico real durante la realización del ejercicio, supone riesgo de sufrir un percance.

- Se otorgará un tiempo prudencial, no más de tres minutos, para que el evaluado se acostumbre al vehículo, haciendo especial hincapié en que adopte en una posición ergonómica y cómoda en el puesto de conducción.

- En la medida de lo posible, el examinador, o auxiliares si los hubiera, tratarán de dar únicamente indicaciones de dirección para trazar el recorrido planificado, absteniéndose de dar cualquier otra explicación o aclaración durante la realización de la prueba.

- Una vez terminado el recorrido establecido se recogerá el disco-diagrama de tacógrafo y se terminará de cumplimentar la hoja de evaluación de la EPC. En este sentido la instrucción general debe ser la obligación de tener la hoja cumplimentada al completo antes de iniciar la evaluación del siguiente candidato.

- Al término de las evaluaciones de todos los candidatos, se comprobará que no falta ninguna hoja de evaluación.

3.2.2 Normas de aplicación

Estas indicaciones se darán en conjunto a todos los participantes en el proceso de evaluación y siempre antes de iniciarse la prueba práctica.

1º Se realizará una presentación personal del responsable de la prueba, quien deberá presentarse como *Examinador* o *Evaluador*, nunca como *Monitor* o *Formador*. En ese momento, si no se hubiera hecho antes, se dará a los candidatos una explicación del motivo por el que se realiza la EPC, señalando la importancia de hacerlo bien y la significación que su resultado tendrá en el conjunto del proceso.

2º La primera indicación general del evaluador será:

Tengan, por favor, preparado su permiso de conducir y su tarjeta CAP, aquellos que la requieran. Si no tienen con ustedes esta documentación deben reunirla ya que sin ella no podrán realizar la prueba. Gracias

3º A continuación se añadirá:

Esta será una prueba en la que se evaluará su habilidad para la conducción de vehículos autocares. Cada uno de los candidatos dispondrá de un tiempo entre treinta y cuarenta y cinco minutos para la realización de un recorrido que yo les iré indicando. Ahora bajaremos al vehículo y les daré otras instrucciones. Les informo de que únicamente podrán subir al autobús aquellas personas participantes en el proceso de evaluación.

4º Una vez en el vehículo se comprobará la validez de los documentos que aporten los participantes; se dirá:

Ahora, vamos a subir al autobús en el que realizaremos la prueba. Entréguenme, por favor, su documentación.

La revisión de los permisos, especialmente las fechas de expedición, vigencia y condiciones restrictivas, permite comprobar que todos están habilitados para realizar la prueba y decidir qué aspirante pasará primero, conforme a su experiencia previa. La recomendación general es que la evaluación inicial se haga al candidato que tenga más experiencia, prioritariamente en autobuses.

5º La última indicación general hará referencia a las instrucciones que se darán durante la prueba; se hará la siguiente observación:

Recuerden que todas las indicaciones que les daré mientras realicen la prueba serán legales y no supondrán, en ningún caso, infringir ninguna norma de circulación.

6º Se nombrará al candidato que se haya elegido para ser el primero para realizar la prueba.

Tras estas indicaciones generales, se procederá a dar instrucciones individualizadas a cada uno de los candidatos, según vayan realizando la prueba, justo antes de iniciar el recorrido:

1º Se le preguntará específicamente por su experiencia profesional con vehículo pesado, comenzando por la posible experiencia con autocares. En un espacio en blanco de la hoja de evaluación de la prueba se tomará nota de los años de experiencia con autocares y con camiones.

2º A continuación se le indicará:

Ahora, tome asiento al volante. Le entrego un disco-diagrama de tacógrafo para que lo cumplimente y lo utilice durante la prueba. Si no conoce su funcionamiento o no está seguro de su instalación, por favor, ponga su nombre y yo le guiaré en su cumplimentación y uso.

3º Una vez instalado el disco, se sugerirá lo siguiente:

Antes de iniciar la marcha, asegúrese de estar en una posición cómoda y ergonómica, verifique que llega bien a los pedales y que visualiza correctamente los espejos retrovisores.

Se añadirá:

Si tiene alguna duda sobre los mecanismos de accionamiento del salpicadero o del display, éste es el momento de resolverla.

4º Una vez resueltas las dudas que pudiera haber, en caso de que el vehículo esté provisto de cinturón de seguridad y el candidato no haya hecho aún uso de él, se le indicará:

Le recuerdo la obligatoriedad de uso del cinturón de seguridad. Por su seguridad, le recomiendo que lo utilice.

5º Finalmente se dirá:

Si no tiene ninguna duda más, arranque el vehículo y siga mis instrucciones durante toda la prueba. Yo le iré indicando, con tiempo suficiente, las direcciones que debe tomar.

Y se añadirá para todos los participantes en general:

A partir de este momento, les pediría que, por nuestra seguridad, si tienen alguna duda o pregunta las hicieran una vez terminada la prueba. Gracias.

En este punto comienza la evaluación, como tal. Durante la duración de la prueba, se debe tratar de mantener una actitud profesional, conservando la

distancia con los evaluados, centrada en recabar la mayor cantidad de evidencias conductuales que permitan sostener la puntuación otorgada.

Por otro lado, únicamente se establecerá contacto con el candidato que esté conduciendo para darle indicaciones de dirección. Éstas deben ser lo más claras posible. Se deben evitar indicaciones ambiguas, que puedan llevar a error o confusión, como por ejemplo *gire a la izquierda*. La recomendación a este respecto es que las indicaciones se den con tiempo suficiente y haciendo referencia a los paneles de preseñalización, de dirección, señales y marcas viales, describiendo la dirección de interés tal y como va a aparecer en el panel informativo.

De este modo controlamos la claridad de la información, evitamos errores de interpretación y, al mismo tiempo, podemos obtener evidencia de la interpretación de las señales, recogida en el ítem número 9 *Atención a la señalización vial*.

El tipo de indicación que debería darse se puede ver en el siguiente ejemplo (Fig. 13):

En la próxima intersección llegaremos a una rotonda. Quiero que tome dirección Guadarrama. Dirección Guadarrama.

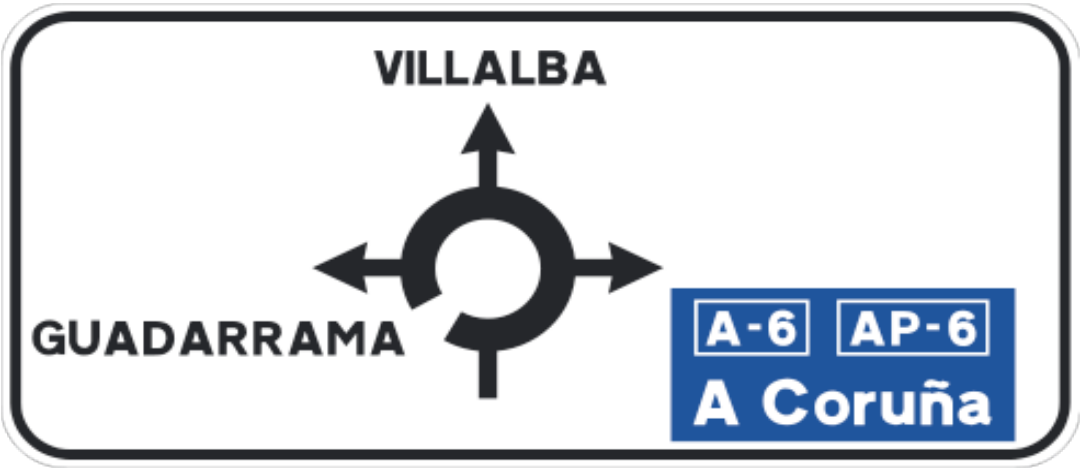


Fig. 13: Ejemplo de indicación de dirección durante la prueba.

En caso de ser necesario repetiríamos la dirección para aclarar cualquier duda. Evitaremos referencias al número de la salida, o indicaciones vagas y genéricas del tipo *giraremos a la izquierda* o *tome la siguiente intersección a la derecha*.

Puede ocurrir que, en alguna ocasión, el candidato se pase de la salida indicada o se equivoque en una dirección. Por mantener el criterio de seguridad en

la prueba, no haremos ninguna indicación ni trataremos de corregir en el último momento, le dejaremos continuar la maniobra y haremos mención del hecho para que tenga constancia de su error.

Si esta equivocación se diera por causa achacable al evaluador, por ejemplo por haber transmitido con retraso la indicación, no se tendrá en cuenta y se dará una explicación del tipo:

No se preocupe, ha sido un error mío. Continúe, le daré indicaciones de dirección más adelante para volver a nuestra ruta.

La interacción con el sujeto evaluado se limitará, exclusivamente, a este tipo de intercambios y nada más. Únicamente se intervendrá en el desarrollo de la prueba para salvar situaciones que sean potencialmente peligrosas para la seguridad del vehículo, sus ocupantes u otros usuarios de la vía, tanto otros vehículos como peatones.

Cuando se haya completado el itinerario marcado, se le dará indicación al candidato de que detenga el vehículo en lugar seguro para el autobús y donde no se interrumpa el tráfico. Se hará especial hincapié en que el autobús se sitúe de tal forma que permita al siguiente candidato iniciar su marcha sin tener que hacer maniobra marcha atrás.

La instrucción final, una vez el vehículo esté estacionado ha de ser:

Por favor, ponga el freno de mano, extraiga el disco del tacógrafo y proceda a cerrarlo.

En ese momento se debe aprovechar para cerrar las últimas puntuaciones de la prueba que restaran por poner. La hoja de evaluación debería estar cumplimentada al menos al 90% antes de que el siguiente candidato comience su prueba.

Al concluir la prueba el último de los candidatos, con el motor apagado y el vehículo estacionado, se puede completar la información del ítem número 11 *Observación del cuadro de mandos*. Para ello se procederá a plantear, uno por uno a los evaluados, preguntas del tipo siguiente:

¿Podría indicarme, por favor, el nivel de combustible con el que circulábamos?

Cualquier otra pregunta es válida, siempre que haga referencia a información proporcionada por el display del salpicadero del autocar.

Posteriormente, se hará, en común para todos los participantes, la siguiente valoración final:

> *Hemos concluido la prueba de evaluación. Muchas gracias a todos por su participación. Deben tener en cuenta que esta prueba no dictamina si se sabe o no conducir; valora el grado en el que se poseen, o no, determinadas habilidades consideradas clave a la hora de conducir autobuses.*

Tras esta indicación general se procederá, uno por uno, a determinar la puntuación final obtenida según los criterios que se describen a continuación en el siguiente punto.

3.2.3 Puntuación

En toda la prueba la puntuación directa es igual a la suma de los totales de cada uno de los trece ítems. La escala de cada uno de los ítems es de 0 a 3, por lo que la puntuación directa más alta posible es 39 puntos.

Se considerará que un candidato tiene un nivel de conducción apto para el ejercicio profesional, cuando obtenga una puntuación total igual o superior a **23 puntos**. Esto supone alcanzar, como mínimo, un 60% de la puntuación máxima. Para pruebas realizadas en vehículos de transporte urbano, la puntuación de corte será de **18 puntos**. Esta puntuación también se corresponde con alcanzar, como mínimo, un 60% de la puntuación. Recuérdese que se eliminan tres ítems del total de la prueba, lo que reduce el máximo de puntos alcanzable.

Los tres criterios establecidos (Conducción económica, Seguridad y Conducción técnica) se puntúan por separado, pero no tienen consideración independiente. Esto implica que se pueda alcanzar el mínimo total exigido aún no llegando al mínimo en alguno de los criterios. El estudio de este aspecto es una de las hipótesis que se plantean para este trabajo de investigación.

3.3 Descripción de otras pruebas utilizadas en el estudio

3.3.1 Para el estudio sobre Atención

En el ámbito del estudio sobre la relación del nivel atencional del candidato con los resultados de la prueba se utilizó el test de atención d2 (Brickenkamp, 2009)

El test de atención d2 es una prueba de tiempo limitado para medir atención selectiva. Mide la velocidad de procesamiento, el cumplimiento de las instrucciones y la precisión en la ejecución de la tarea. Permite estimar la capacidad atencional y de concentración en personas de 8 a 60 años de edad.

Se aplica de manera colectiva, con un tiempo estimado de aplicación de entre 8 y 10 minutos, incluyendo las instrucciones. A pesar de ser un test relativamente antiguo en su construcción está aún muy vigente y está siendo utilizado con frecuencia en investigaciones relacionadas con la medida de la atención (Brickenkamp, 2009; Jiménez, et al., 2012)

El área de trabajo está formada por 14 líneas con 47 caracteres (658 elementos) Estos estímulos pueden contener las letras “d” o “p”, acompañadas con una, dos, tres o cuatro pequeñas líneas verticales situadas en la parte superior o en la parte inferior de la letra. Para cada línea el sujeto dispone de 20 segundos, el tiempo real efectivo para del candidato para realizar la prueba es, por tanto, de 4 minutos 40 segundos.

La tarea consiste en revisar de izquierda a derecha el contenido de cada línea y marcar toda letra "d" que vaya acompañada de dos líneas, bien en pareja en la parte superior o en la parte inferior, bien una arriba y otra debajo de la letra.

Las puntuaciones que se obtienen como resultado se presentan a continuación:

- TR: Total de respuestas. Es el número de elementos intentados. Es una medida muy fiable de la atención (sostenida y selectiva) así como de la velocidad de procesamiento.
- TA: Total de aciertos. Es el número de elementos relevantes que se han marcado correctamente. Mide la cantidad de trabajo que se puede realizar con efectividad, es decir la precisión en la ejecución.
- O: Omisiones. Es el primer tipo de error que se puede cometer. Implica dejar sin marcar aquellos elementos relevantes que deberían haber sido marcados.
- C: Comisiones. Es el segundo tipo de error. En este caso supone haber marcado elementos no relevantes, es decir marcar letras que no deberían haber sido marcadas.
- TOT: Total de elementos procesados. Es una medida de la cantidad de trabajo realizado tras haber eliminado el efecto de los errores (O+C), una medida de efectividad en la prueba.
- CON: Índice de concentración. Medido a partir de la diferencia entre el total de aciertos TA y en número de comisiones C, esto es, por tanto, una medida de la precisión en la ejecución de la prueba.

3.3.2 Para el estudio sobre personalidad

En la presente investigación se utilizó, como método de estudio del perfil de personalidad de los participantes, el test NEO PI-R (McCrae y Costa, 1999). De sus resultados se estudiaron únicamente los cinco grandes factores, sin entrar en describir las puntuaciones en sus distintas facetas. Para ello se utilizó la versión

reducida del test, la cual consta de 60 preguntas que evalúan los siguientes rasgos generales de personalidad:

- Neuroticismo (N): La escala N del NEO PI-R mide una dimensión clásica en los estudios sobre personalidad. En este caso la escala es una medida de la personalidad normal y no se considera una medida psicopatológica. Sin embargo, sigue siendo una evaluación fiable de la estabilidad emocional del sujeto y de su capacidad de enfrentamiento a situaciones estresantes cuando la medida en este factor es baja.

- Extraversión (E): La sociabilidad y la expansión social no es el único aspecto que influye en este factor. La asertividad y la búsqueda de actividades estimulantes también caracterizan al tipo de personalidad extravertido.

- Apertura (O, Openness): Esta escala, por menos conocida y poco frecuente en los estudios precedentes, es la más estudiada en la última época. El perfil de personalidad que muestra apertura es el de una persona interesada por la experiencia, con interés por la integración activa y por la sensibilidad estética. Acepta nuevas ideas y valores considerados no convencionales. Por contra, el excesivo convencionalismo y una conducta en apariencia conservadora caracterizan a la puntuación baja en este factor, no como una reacción defensiva generalizada, si no, simplemente, como muestra de un campo de interés más reducido. El cuestionamiento de la autoridad, propio de los niveles altos en este factor, es un elemento a tener en cuenta en relación con la conducción de vehículos.

- Amabilidad (A): Representa, al igual que la Extraversión, una dimensión interpersonal. El polo de Amabilidad alta se caracteriza por una capacidad elevada de mostrar conductas altruistas y de simpatizar con el resto. La capacidad de cooperación y empatía es una facultad muy apreciable en el entorno de la conducción.

- Responsabilidad (C, Concientiousness): Por último, el autocontrol, entendido como un proceso activo en el que las tareas se planean y ejecutan con organización, es la base de una puntuación alta en este

factor. La escrupulosidad, la puntualidad y la fiabilidad caracterizan a los perfiles con puntuaciones altas en este factor.

Siguiendo el esquema general de trabajo que se ha mantenido hasta ahora, el NEO PI-R se aplicó a una parte de la muestra general de sujetos participantes en el estudio de propiedades psicométricas del cuestionario de evaluación de conducción. De esta forma, se ha podido realizar un estudio de las posibles relaciones entre las puntuaciones en la prueba de personalidad y la evaluación recibida en el test práctico de conducción.

3.4 Planteamiento de hipótesis

3.4.1 Estado de la cuestión

El sector del transporte de viajeros en España es un sector fuertemente intervenido, con una amplia supervisión y control por parte de las diferentes administraciones, que utilizan diferentes fórmulas de gestión para cumplir con su obligación de Servicio Público; desde las Autorizaciones de transporte hasta las Concesiones Administrativas, más restrictivas estas últimas, exclusivas de los transportes públicos de carácter regular, que las primeras.

Es, además, un sector atomizado. En abril del año 2010, según el último Plan de Actuaciones en el sector del Transporte público de viajeros en Autobús (Plan PLATA), en España había 4.533 empresas autorizadas para el transporte de viajeros (3.732 para servicio público y 801 para servicio privado). El tamaño medio de las empresas ha venido creciendo desde el año 1999, hasta alcanzar actualmente los diez vehículos por empresa. A pesar de que ese dato acredita la tendencia actual de ir hacia la concentración empresarial, aún hoy la gran mayoría de las empresas autorizadas no superan los cinco vehículos (OSTC, 2010).

Por otro lado, la edad media de ese parque móvil es de once años y también cabe señalar que únicamente un 3% de los trabajadores de este sector son autónomos, destacando la figura del trabajador asalariado y de los contratos temporales, principalmente para los periodos escolares (DGTT, 2011b).

Un último dato que también muestra claramente la atomización del sector es la existencia de 55 convenios colectivos en vigor, un número mayor que el total de

provincias españolas y que dificulta, en muchas ocasiones, la adopción de medidas comunes en beneficio del sector (DGTT, 2011a).

La mayoría de estas pequeñas empresas subsisten dando servicio a autorizaciones de transporte en líneas regulares de uso especial, generalmente colegios y empresas, o realizando una intensa actividad de transporte de carácter discrecional, la cual está disminuyendo paulatinamente en los últimos tiempos.

Así las cosas, frecuentemente se ha venido sosteniendo que la ejecución de procesos de selección estandarizados con pruebas y evaluaciones específicas para el puesto de conductor era algo casi irrealizable en un sector tan atomizado como el del Transporte de Viajeros, donde las empresas destinan pocos recursos a esa faceta.

El perfil tradicional y familiar de la gran mayoría de las más de 4.500 empresas que forman el entramado empresarial del transporte de viajeros en España, ha convertido a la selección de conductores en un proceso de conocimiento personal y de aprovechamiento de las relaciones, dejando de lado la profesionalización de un proceso que debería ser medible y con criterios de aplicación común. La falta de definición clara de un perfil y la escasa profesionalización de los departamentos de Recursos Humanos en el sector, mayoritariamente relegados a la mera función de administración de personal, carga de subjetividad y de conocimiento intuitivo a este tipo de procesos de selección.

Mi experiencia personal en el sector es que, en la mayoría de las ocasiones, es personal propio de la empresa, no especializado, el que realiza la tarea de la selección, sin más formación ni conocimiento sobre selección de personal que el que puede aportar la experiencia de muchos años como conductor o como jefe de taller, pero sin una preparación específica para una tarea de gestión de Recursos Humanos como ésta. Se evidencia, en este punto, una falta de especialización en los responsables de realizar este trabajo que responde, más bien, al criterio tradicionalmente extendido y arbitrario de que un conductor experimentado es el más dotado para evaluar a futuros conductores. Esta visión tampoco favorece que los procedimientos de selección se puedan estandarizar y aplicar con rigor.

Por otra parte, tradicionalmente, el sector del transporte viene siendo un entorno en el que la rotación de personal es bastante significativa, presentando una alta temporalidad con únicamente un **12%** de contratos indefinidos. Hasta el año 2007, según datos del Observatorio Social del Transporte por Carretera 2009 (OSTC,

2011), las ofertas de empleo para conductores profesionales (camión y autobús) mostraban un significativo incremento interanual, así como también lo han venido haciendo las demandas y solicitudes de empleo para trabajar como conductor, indicando que, aunque en los dos últimos años esa tendencia se ha visto frenada por la situación económica, la alta rotación de trabajadores en este sector es una situación habitual; algunas de las principales empresas de transporte del país reportan unos índices de rotación cercanos al **35%** de la plantilla anual. Los cambios de empresa así como de modo de transporte, sobre todo del sector de viajeros hacia mercancías, han sido habituales para los conductores profesionales en España. Las mejoras en las condiciones salariales, tipo de trabajo o mejoras en la calidad de vida ofertada, han propiciado desde siempre los cambios de empresa en un sector donde, además, la selección de personal está prácticamente ausente. Esto facilita aún más la decisión de dejar un trabajo ante la práctica seguridad de que, sin más ayuda que la experiencia previa, en un plazo breve y sin demasiado esfuerzo, se podrá acceder a un nuevo puesto de trabajo.

Así las cosas, la posibilidad de aportar valor añadido al proceso de incorporación de los trabajadores de este sector, vía el establecimiento de un procedimiento estandarizado de selección, aparece como una seria opción de mejora, que, como mínimo, dotaría de objetividad y de validez aparente al acceso a los puestos de trabajo de conductor profesional.

Para ello debe partirse del establecimiento de una definición clara y comúnmente aceptada de la profesión. En este caso, donde el perfil de interés es el de Conductor de Autobús, se ha tomado la definición de la competencia general, contenida en la especialidad formativa *Conductor de Autobús*, código TMVC40 del Catálogo de Especialidades Formativas del Servicio Público de Empleo Estatal, y que define el perfil profesional del conductor de autobús de la siguiente forma:

> [...] efectúa el transporte de viajeros por carretera de manera segura, responsable y económica, respetando las normas en vigor y las instrucciones/programa de servicio; atiende e informa adecuadamente a los pasajeros. Prepara y aplica el plan de mantenimiento preventivo del autobús y equipos auxiliares, reparando en caso necesario posibles averías o disfunciones simples. Actúa en caso de accidentes o siniestros de acuerdo con los procedimientos definidos.

Esta definición de la competencia general aporta todos aquellos aspectos que debe conocer y aplicar un conductor profesional de autobús y, por lo tanto, aquellas facetas susceptibles de ser evaluadas y que deben ser tenidas en cuenta a

la hora de desarrollar un procedimiento de selección objetivo para conductores de autobús.

Intuitivamente, la conducción es considerada la actividad principal de un trabajador profesional del sector de transporte por carretera. Sin embargo, la propia descripción de la competencia general permite valorar la importancia que otras actividades, como la atención al público o los conocimientos mínimos de mecánica, tienen en el correcto ejercicio profesional.

Obviamente estos aspectos deberán ser tenidos en cuenta para establecer aquellos puntos que formen parte de un proceso de selección completo y que ambicione cubrir todas las facetas definidas en la descripción de la competencia general. Sin embargo, como primer aspecto fundamental, debería establecerse una prueba de evaluación que permitiera estimar, de una forma objetiva y con los necesarios niveles de fiabilidad, en qué grado un aspirante a un puesto de trabajo como conductor cumple con los conocimientos y habilidades mínimos para realizar de manera *segura, responsable y económica* su labor.

Esta prueba de evaluación debería testar estos puntos de una manera fiable, con una serie de criterios comunes de aplicación, conocidos, probados y contrastados, que hagan que su uso sea sencillo e intuitivo y con la necesaria independencia del evaluador, de manera que asegure su objetividad en la medida.

Por la utilidad que inicialmente se le supone a una prueba como la descrita y por no existir en el mercado español nada similar a disposición de las empresas interesadas en realizar procesos de selección de conductores de una forma profesional y objetiva, parece lógico pensar que el desarrollo de una herramienta de estas características es necesario y que sería recibido como un avance importante y beneficioso para la profesión y el sector.

> *¿Es entonces posible establecer las bases de un procedimiento estandarizado, de uso común y de utilidad para la selección de personal de conducción en las empresas de transporte, a partir de la elaboración de una prueba de evaluación de las habilidades prácticas de conducción de un candidato?*

El desarrollo de la presente tesis doctoral trata de dar respuesta a esta pregunta elaborando una prueba de evaluación nueva, que siga los criterios marcados por la definición del perfil de la profesión y que sea analizada de manera

exhaustiva para poder describir con precisión su estructura interna y sus propiedades psicométricas.

3.4.2 Objetivos del estudio

El objetivo principal de este trabajo de tesis es elaborar y poner en funcionamiento una prueba práctica de evaluación, elaborada para conductores de autobús, que evalúe las aptitudes y habilidades que son requisito específico para realizar una conducción segura y económicamente eficiente, tal y como recoge la definición de la competencia general del perfil profesional del conductor de autocar y que pueda ser utilizada en procesos de selección.

Además, se plantean una serie de objetivos secundarios, los cuales se relacionan a continuación:

(I) Estudiar la Validez del test aportando evidencias de fuentes de validez basadas en su contenido, en su estructura interna y sus relaciones con otras variables.

(II) Analizar la estructura dimensional de la prueba, para asegurar que puede hablarse de ella como un test unidimensional.

(III) Estudiar la fiabilidad de la prueba en términos de consistencia interna.

(IV) Realizar un análisis de las propiedades de los ítems, tanto desde el punto de vista de la Teoría Clásica, como de la Teoría de Respuesta al Ítem.

(V) Investigar las relaciones que puedan existir entre la variable *Edad* y los resultados de la prueba.

(VI) Estudiar los resultados de la evaluación realizada con el test en relación con la *Experiencia* previa de conducción que aporten los candidatos.

(VII) Estudiar las relaciones entre la variable *Sexo* y los resultados de la prueba, verificando si puede considerarse la existencia de sesgo dependiendo del género del/la candidato/a.

(VIII) Investigar la relación de la variable *Atención* con los resultados de la prueba.

(IX) Estudiar en qué forma afectan a los resultados de la prueba diferentes rasgos de la variable *Personalidad*.

Por otra parte, como consecuencia natural del trabajo realizado un último objetivo está relacionado con el desarrollo de un manual de aplicación que describa el uso de la herramienta, estableciendo claramente los criterios observacionales y las evidencias conductuales a las que debe prestarse atención y que deben ser recogidas, para establecer una puntuación objetiva. De esta manera cualquier evaluador que utilice la herramienta podrá conocer perfectamente su uso y aplicación. Éste es un criterio clave para que la fiabilidad interjueces de la prueba sea alta y contribuya a la objetividad de las evaluaciones realizadas con ella.

En definitiva, se trata aquí de sentar las bases de una herramienta calibrada y validada que pueda servir para un uso generalizado. No solo para evaluar a conductores profesionales en procesos de selección, con vistas a posibles incorporaciones en empresas del sector, sino que también pueda ser utilizada para otro tipo de estudios sobre Seguridad Vial, como por ejemplo el efecto de un determinado tipo de formación sobre las habilidades de un conductor o el efecto de la experiencia sobre la seguridad en carretera.

3.4.3 Hipótesis de trabajo

Para cubrir los objetivos planteados se definen diversas hipótesis que serán comprobadas en el estudio que se presenta en esta tesis doctoral.

(I) Si el test ha sido construido correctamente se podrá verificar una importante cantidad de evidencias de su validez, basadas en su contenido (representación de todo el dominio relacionado con el constructo, jueces expertos que determinen la relación entre el test y el constructo)

(II) Se hipotetiza que, si el test está bien construido su estructura interna será unidimensional, respondiendo así a la representación del constructo evaluado, *habilidad en la conducción*.

(III) Si se comprueba que la estructura del test es congruente con una estructura unidimensional, su fiabilidad, en términos de consistencia interna, será alta.

(IV) Las propiedades de medida de los ítems, en términos de dificultad y discriminación, no serán diferentes comparando su cálculo con TCT frente a TRI.

(V) Si la prueba evalúa exclusivamente habilidad en la conducción, se espera que exista una relación con respecto a la variable *Edad*, de tal forma que los candidatos de mayor edad obtendrían mejores resultados.

(VI) También, por medir habilidad en la conducción, es esperable que se compruebe una influencia de la variable *Experiencia* en los resultados de la prueba. Así, se hipotetiza que los sujetos con más experiencia con vehículo industrial obtendrán mejores resultados en el global de la prueba.

(VII) No es esperable que la prueba esté sesgada en función del sexo de los participantes, por lo que no se esperan diferencias significativas entre las puntuaciones de hombres y mujeres.

(VIII) La variable *Atención* está relacionada con seguridad en la conducción y con un menor índice de siniestros, pero no existe una relación clara con la *habilidad en la conducción*, la cual se evalúa con esta prueba, es por ello que no prevén diferencias significativas en las puntuaciones dependiendo de las capacidades atencionales del conductor.

(IX) Del mismo modo, diferentes factores y tipos de personalidad se relacionan con seguridad en la conducción y un menor número de accidentes, pero no directamente con la variable *habilidad en la conducción*, por lo tanto tampoco es esperable que aparezcan diferencias en las puntuaciones totales del test dependiendo de diferentes perfiles de personalidad.

Un hombre con un reloj siempre sabe qué hora es, un hombre con dos relojes nunca está seguro.

Ley de Segal,descrita en las Leyes de Murphy

Propiedades Psicométricas de la Prueba de Evaluación: Fiabilidad y Validez

4

Se describen en este capítulo todos los detalles relativos a la formación de la muestra. Se aportan resultados de su descripción general, periodo de recogida de datos y tratamiento de la información.

Se presentan los resultados del estudio de validez realizado, con especial mención a la prueba de evidencias basadas en el contenido y de estructura interna del test, detallando la técnica estadística utilizada para estimarla. Se aportan también datos del estudio de otras pruebas complementarias realizadas.

Se detallan los datos relativos al estudio de fiabilidad realizado, especificando los fundamentos metodológicos de la técnica y la información que aporta.

También se exponen en este capítulo los resultados del análisis de los ítems realizado desde dos perspectivas diferentes, pero complementarias. Se trazan las líneas generales de la técnica desde la Teoría Clásica de los Test y también se precisan detalles desde la Teoría de Respuesta al Ítem, justificando el uso de ambas técnicas y la visión conjunta no excluyente que ambas aportan.

Por último, se describe y justifica el procedimiento seguido para el cálculo de un baremo general de la prueba

4.1 Descripción general de la muestra

Para la realización de los estudios incluidos en esta tesis doctoral se recogieron datos en el periodo comprendido entre julio de 2007, momento a partir del cual el uso de la prueba de evaluación fue aprobado y comenzó a aplicarse en todas las pruebas de selección, y mayo de 2011.

Todos los datos correspondientes a las pruebas de evaluación son mecanizados e incluidos en la base de datos general con la que se trabaja en el departamento de Selección y Formación de la empresa. Junto con la información específica de la evaluación de conducción se incluyen otra serie de datos personales, que permiten la identificación del candidato, y profesionales que son posteriormente utilizados para la elaboración de informes por parte del personal propio del departamento.

Para este estudio se exportaron los datos relativos a las pruebas de conducción junto con datos demográficos (género, edad, provincia, antigüedad del permiso) y de experiencia laboral. Estos datos se utilizaron exclusivamente a efectos de estudio estadístico preservando, en todo momento, el anonimato de los sujetos.

Durante el periodo estudiado, de casi cuatro años, se recogieron datos sobre **4.425** pruebas de evaluación, realizadas con el test aquí estudiado. Todos los participantes cumplían el requisito inicial de estar en posesión de un permiso de conducción de la clase D en vigor. Los datos provienen de pruebas realizadas para alguno de estos tres propósitos:

1. **Pruebas de selección de personal conductor**: Cuando un conductor, con más o menos experiencia, se postula como

candidato para trabajar en la empresa pasa previamente por una prueba de selección de personal. Esta prueba, además de diferentes test psicológicos y entrevistas personales incluye una evaluación del nivel de conducción, realizada con el test aquí descrito.

2. **Evaluaciones realizadas a conductores de la empresa en activo**: Cuando la empresa así lo requiere conductores en activo pasan por una evaluación de su nivel de conducción. Esta evaluación es realizada por evaluadores expertos en la utilización del test.

3. **Registros de evaluación tras pasar por alguna sesión o curso de formación**: Aquellos conductores, aspirantes o trabajadores en activo, que participan en alguna de las acciones formativas sobre conducción que realiza el departamento de Selección y Formación, pasan así mismo por una evaluación realizada con el test al finalizar la acción formativa.

En cualquiera de estas tres situaciones se genera un registro que relaciona a cada participante y evaluador con los resultados de la prueba realizada.

Para el estudio de las propiedades psicométricas del test, se utilizaron únicamente aquellas pruebas que tenían puntuaciones en los trece ítems que conforman el test completo. Se excluyeron todas las evaluaciones en las que había respuestas en blanco en alguno de los ítems. Los motivos más frecuentes por los que sucede esto son, o bien por haberse realizado la prueba con un vehículo automático[5], o bien por errores en el proceso de transcripción de los datos.

Tras haber depurado la tabla de datos conforme a este criterio y habiendo eliminado las puntuaciones outlier[6] la muestra definitiva con la que se realizaron los estudios generales sobre las propiedades de la prueba quedó reducida a **3.857** pruebas.

[5] Cuando la prueba se realiza con un vehículo automático los ítems relacionados con el uso del cambio y control de las revoluciones no pueden ser registrados, por motivos obvios. Además, en general, este tipo de modelos carece también de retardador hidráulico de frenada, por lo que cuando se utiliza un vehículo automático, en la mayoría de las ocasiones, tampoco se puede registrar ningún dato relacionado con el uso del retarder.

[6] Las puntuaciones outlier son aquellas que resultan sospechosas dentro de la distribución normal de las puntuaciones. En este caso cualquier puntuación por encima de 3 en cualquiera de los ítems se consideró extraña y fue eliminada dado que la puntuación de cada uno de los ítems solo puede estar entre 0 y 3.

Para este total de pruebas, el intervalo de puntuaciones está en un rango de 3 a 39 puntos (Tabla 5), con una media de 22,25 y una desviación típica de 6,36. La mediana de la muestra se sitúa en 23 puntos, puntuación que coincide con la definida por los creadores de la prueba como la mínima para obtener una evaluación de apto en la evaluación. Los resultados pueden observarse en la tabla que se muestra a continuación.

Estadísticos

Puntuación Total Test		
N	Válidos	**3857**
	Perdidos	**0**
Media		**22,25**
Mediana		**23,00**
Desv. típ.		**6,357**
Mínimo		**3**
Máximo		**39**
Percentiles	25	**18,00**
	50	**23,00**
	75	**27,00**

Tabla 5: Estadísticos descriptivos de la puntuación del test para la muestra general

Se ha calculado un error máximo de medida de 0,20 al nivel de confianza del 95%.

Todas estas pruebas fueron supervisadas por 35 evaluadores diferentes. Todos habían pasado previamente por un periodo de formación de tres días durante los cuales se les enseñó el uso correcto de la herramienta con sesiones teóricas y prácticas; además, se les entregó a todos una copia del manual de aplicación de la prueba, que fue elaborado específicamente para esa formación.

Aún así, más del 80% de las pruebas (82,50%) fueron realizadas por 6 evaluadores. Estos, miembros permanentes del departamento de Selección y Formación, son los que con mayor frecuencia realizan pruebas y utilizan la herramienta de evaluación.

En la Figura 14 puede verse la distribución de las puntuaciones obtenidas en el test.

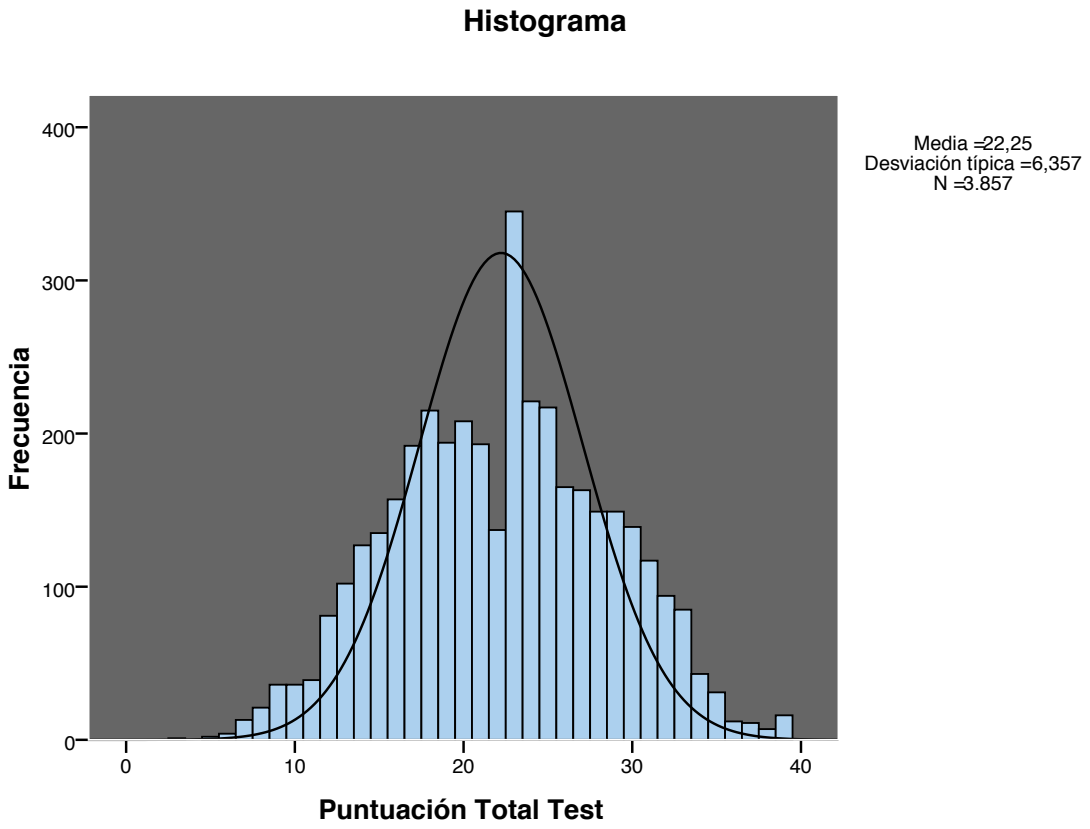


Fig. 14: Histograma de distribución de la puntuación total en el test

El 66% de las puntuaciones se sitúa dentro del rango de mas menos una desviación típica; no alcanza el 68% que marca la norma (Agresti y Finlay, 2009), pero sí que visualmente parece distribuirse de acuerdo con la curva normal. Sin embargo, el estudio de ajuste a la normalidad realizado mediante la prueba de Kolmogorov-Smirnov indica que la muestra no se distribuye normalmente, como muestran los datos de la salida del análisis realizado con la función ksnormTest() implementada en el paquete fBasics de R (R_Core_Team, 2013).

One-sample Kolmogorov-Smirnov test

Test Results:

STATISTIC:

D: 0.9709

P VALUE:

Alternative Two-Sided: < 2.2e-16

Alternative Less: < 2.2e-16

Alternative Greater: 1

En cuanto a la distribución de la muestra, de acuerdo con el sexo de los participantes, la mayoría de los evaluados eran hombres, un **84,30%**, mientras que un **13,50%** eran mujeres. Hay un **2,2%** de las pruebas recogidas de las que no se tiene registro del género del participante. En total, 3.251 participantes eran hombres y 522 mujeres. De 84 pruebas no se recogieron datos a este respecto.

El rango de edad para esta muestra se sitúa entre los 21 y 64 años. El límite inferior viene marcado por la normativa en materia de Tráfico y Seguridad Vial. Conviene recordar que, hasta diciembre de 2009, la edad mínima exigida para la obtención de un permiso de la clase D era de 21 años. A partir de esa fecha, con la entrada en vigor del nuevo Reglamento General de Conductores, la edad mínima se elevó a 24 años para la obtención de un permiso que habilite administrativamente para conducir autobuses (ver tabla 4, p. 29). De este modo, aunque la normativa actual limita el acceso administrativo a este permiso hasta no cumplir los 24 años, cuando se inició la recogida de datos existía la posibilidad legal de estar en posesión de un permiso de autobús con una edad inferior a la referida.

El límite superior del rango de edad está sujeto, por razones obvias, a la edad máxima de jubilación en una empresa. Actualmente no existe límite de edad para la renovación de un permiso clase D; solo se exige superar el reconocimiento médico psicotécnico obligatorio. Sin embargo, la edad de jubilación en España impone que una empresa no tenga en su nómina a trabajadores con edad superior a los 65 años.

La media de edad de la muestra es cercana a los 38 años (**M=37,89**) con una desviación típica superior a 8 (**DT=8,53**) (Tabla 6).

Tramos de Edad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Menores de 25	**240**	**6,2**	**6,6**	**6,6**
	de 25 a 30	**555**	**14,4**	**15,4**	**22,0**
	de 30 a 35	**744**	**19,3**	**20,6**	**42,6**
	de 35 a 40	**707**	**18,3**	**19,6**	**62,1**
	de 40 a 45	**638**	**16,5**	**17,6**	**79,8**
	de 45 a 50	**440**	**11,4**	**12,2**	**92,0**
	Mayores de 50	**291**	**7,5**	**8,0**	**100,0**
	Total	**3615**	**93,7**	**100,0**	
Perdidos	Sistema	**242**	**6,3**		
Total		**3857**	**100,0**		

Tabla 6: Distribución de la muestra según la edad

Así mismo, se trató de que la muestra fuera representativa de la población española de conductores. Para ello el muestreo se realizó en toda España,

obteniendo datos de conductores de 48 de las 50 provincias españolas. Únicamente quedaron fuera Teruel y Huelva, áreas donde la empresa no tiene presencia. Del resto de las provincias, la mayor o menor presencia viene marcada por necesidades propias de la distribución geográfica de la empresa. De allí donde tiene mayor presencia y volumen de negocio fueron las áreas de las que se pudieron recabar más datos. Es por esto que Madrid, Asturias, León, Cantabria, Vizcaya y Alicante reúnen casi el 65% del total de la muestra (64,80%).

Otras áreas con presencia relevante de la empresa tienen también un volumen considerable de representación en la muestra. Murcia, Barcelona, Toledo y Zaragoza se acercan al 10% del total de la muestra (Fig. 15)

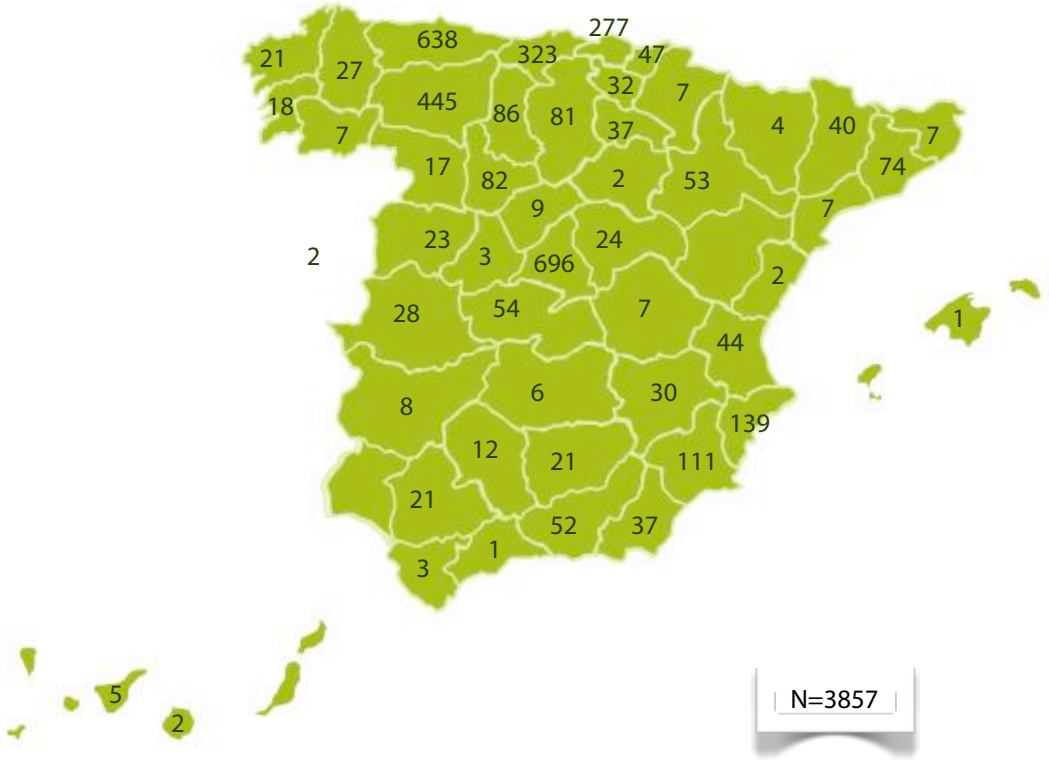


Fig. 15: Distribución del número de participantes por provincia

En el mapa de distribución puede observarse que en la muestra total figuran dos ciudadanos portugueses. Se tomó la decisión de incluirlos en virtud de que su experiencia previa como conductores había sido mayoritariamente en España, como trabajadores de empresas españolas (a pesar de que su residencia no se había fijado en España y se trasladaban todos los fines de semana a Portugal tras trabajar toda la semana en España).

Por último, se recogió también el dato referente a la antigüedad del permiso de conducir del candidato. Esta antigüedad se calculó expresamente a fecha de

realización de la prueba de evaluación. Por lo tanto, expresa los años pasados desde la obtención del permiso de autobús hasta la fecha de realización de la evaluación, ya fuera por acudir a una prueba de selección o por otro motivo.

En este punto conviene señalar que no se tratará de utilizar la antigüedad del permiso como referente de la experiencia profesional de los sujetos. Indudablemente es un indicador, pero no puede confiarse en él como un indicador fiable, toda vez que la obtención del permiso no lleva implícito su uso. Más adelante se describirá un estudio específico realizado para determinar el efecto de la experiencia sobre la prueba. Baste decir aquí que, para efectuarlo, se tuvo en cuenta la experiencia real de un grupo representativo del que se codificó la experiencia como trabajador, incluida en su historia laboral.

En la Tabla 7 puede observarse cómo más del 65% de la muestra realizó la prueba con una antigüedad inferior a 5 años y, en casi el 40% de los casos, esta antigüedad no llegaba al año en el momento de someterse a la evaluación. Esto se explica por la mayor presencia de pruebas de selección que de evaluación de trabajadores en activo. La mayor actividad del departamento y el principal ámbito de uso de la herramienta es el área de selección, donde acuden candidatos con el permiso recién obtenido, buscando una posibilidad de incorporarse a trabajar.

Antigüedad del Permiso

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Un año o menos	**1327**	**34,4**	**39,2**	**39,2**
	de 2 a 5 años	**884**	**22,9**	**26,1**	**65,4**
	de 5 a 10 años	**342**	**8,9**	**10,1**	**75,5**
	de 10 a 15 años	**319**	**8,3**	**9,4**	**84,9**
	de 15 a 20 años	**287**	**7,4**	**8,5**	**93,4**
	más de 20 años	**222**	**5,8**	**6,6**	**100,0**
	Total	**3381**	**87,7**	**100,0**	
Perdidos	Sistema	**476**	**12,3**		
Total		**3857**	**100,0**		

Tabla7: Distribución de la muestra según la antigüedad del permiso

La antigüedad media para esta muestra es superior a 6 años (**M=6,18**), con una desviación típica de más de 7 años (**DT=7,6**).

4.2 Análisis de Validez

La cuestión de la validez de un instrumento de medida no es un tema menor. La American Psychological Association (APA) considera la validez como el aspecto más importante de la evaluación mediante el uso de test (AERA, APA, y NCME, 1999). Indudablemente es necesario que un test mida con fiabilidad, precisión y consistencia en el tiempo, pero ésa no es condición suficiente para establecer su bondad como instrumento de medida. *Además de medir de forma fiable una herramienta de evaluación debe medir aquello que intenta medir* (García-Cueto, 1993).

Las evidencias de validez de la medida es, junto con la fiabilidad, que se tratará más adelante, condición necesaria para valorar un instrumento como una buena herramienta de medida. Existen casi tantas definiciones del concepto de validez, como autores se han interesado por el tema. García Cueto (1993) resume las más relevantes y también aporta la suya propia:

> Un test, como cualquier otro instrumento de medida, es válido si sirve para medir adecuadamente aquello para lo que fue pensado como tal instrumento de medida (García-Cueto, 1993; p. 87)

La definición de validez cobra, entonces, un sentido más filosófico y conceptual que estadístico o psicométrico, aunque en una primera época se consideraba válido cualquier test que correlacionaba con un criterio (Guilford, 1946).

Sin embargo, esta concepción de la validez ha ido evolucionando y desarrollándose hasta el momento actual. Ahora se acepta de manera general la visión de Messick, uno de los investigadores que más ha trabajado e influido en la validez como concepto clave de la medición psicológica tantas veces citada (Muñiz, 1998; Paz, 1996; Pérez-Gil, Chacón Moscoso, y Moreno Rodríguez, 2000), y según la

cual la validación implica la utilización del método habitual en la investigación científica.

> La validación de un test abarca todas las cuestiones experimentales, estadísticas y filosóficas por medio de las cuales se evalúan las hipótesis y teorías científicas. (Messick, 1989; p. 14).

Esto hace que, en la actualidad, la validación de un test se realice como un proceso en el cual se acumulan evidencias que validan las inferencias realizadas a partir de los resultados obtenidos en él en lugar de considerarla como un procedimiento sin más.

> No se trata de medir con precisión, además, y, sobre todo, hay que garantizar que las inferencias y decisiones que se hacen basadas en esas mediciones son correctas. (Muñiz, 1998).

La validez se entiende, entonces, como un concepto unitario que presenta estrategias diferentes de validación dependiendo de las distintas fuentes de evidencias (AERA, et al., 1999):

- Evidencias basadas en el contenido del test
- Evidencias basadas en la estructura interna del test
- Evidencias basadas en las relaciones con otras variables

Estas tres, entre otras, son consideradas como diferentes facetas de un todo. Distintos tipos de evidencia que será necesario recabar para poder determinar que las inferencias realizadas a partir de una prueba pueden ser consideradas válidas. Un proceso de validación ideal debe incluir evidencias obtenidas de muy diferentes formas y fuentes (AERA, et al., 1999).

De este modo, se trata de un proceso abierto, donde siempre cabe añadir nueva experiencia y nuevas pruebas empíricas que corroboren o refuten la pertinencia de las inferencias hechas con el test (Muñiz, 1998).

4.2.1. Evidencias basadas en el contenido del test

La validez de contenido sirve para determinar hasta qué punto los elementos del test son una muestra representativa y adecuada de la población de contenidos posibles relacionados con el campo de interés que está siendo estudiado.

Es decir, informa de en qué grado el test constituye una muestra que recoja todos los aspectos relevantes del objeto de estudio.

Habitualmente, el proceso de consecución de evidencias sobre el contenido del test se realiza a partir de análisis racionales del objeto de estudio. Para esto existen diferentes propuestas (Guion, 1977; Yalow y Popham, 1983), pero han sido Crocker y Algina (1986) quienes mejor han definido la sistemática para ese análisis racional, tal y como describe García Cueto (1993):

1. Definir el campo o dominio de interés del test.
2. Seleccionar a un grupo de *expertos* en el tema o la variable que se vaya a someter a medida.
3. Proporcionar un marco estructurado para el proceso de emparejar los ítems con los temas de interés de medida del test.
4. Seleccionar los ítems adecuados, a través del proceso de emparejamiento ítem-campo de interés.

Un test, de la misma manera que el tacógrafo en un autobús, es un instrumento de medida. Sin embargo, el primero encuentra más dificultades cuando se trata de definir aquello que mide. El concepto, rasgo o teoría psicológicos para cuya medida se construyó el test es una propiedad del sujeto, pero no es directamente observable, se trata de un rasgo latente que solo es deducible a partir de evidencias conductuales.

Las evidencias basadas en el contenido del test pueden ser identificadas en el detalle con que se describe el constructo que se quiere medir y el contenido del test (AERA, et al., 1999).

Estas evidencias, también pueden estar basadas en un análisis, bien lógico, bien empírico, de la corrección con la que el test representa el dominio de

contenido (AERA, et al., 1999). También, pueden incluir juicios de expertos sobre la relación del test, o partes del test, con el constructo.

Las evidencias basadas en el contenido del test persiguen identificar, por un lado, la existencia del constructo hipotetizado, lo cual solo puede lograrse desde un marco teórico y poniéndolo en relación con otros constructos (Paz, 1996). Por otro lado, estas evidencias deberán aportar prueba de que, efectivamente, el test en cuestión es un instrumento de medida adecuado para ese constructo.

Este aspecto de la validez se convierte en pieza clave para todo el proceso de verificación de la inferencias realizadas a partir de los resultados de un test, hasta el punto de que en los actuales standards se le considera como otra evidencia más.

> Evidences based on test content can include logical or empirical analyses [...] of the relevance of the content domain to the proposed interpretation of test scores (AERA, et al., 1999 p. 11)

4.2.2 Evidencias basadas en la estructura interna del test

Estudiar la estructura interna de un test proporciona muestra de hasta qué punto las relaciones entre los ítems y los factores se ajustan al constructo/s que fundamenta/n las interpretaciones propuestas de sus resultados (AERA, et al., 1999)

En el proceso de determinación de estas evidencias, el Análisis Factorial es un método frecuentemente utilizado. Combina diferentes procedimientos matemáticos, con el objetivo principal de proporcionar, a partir de las correlaciones entre variables, *l*a resolución lineal de un conjunto de variables en términos de (generalmente) un número menor de categorías o **factores** (Harman, 1980)**.**

En esencia, el Análisis Factorial es un método multivariante que puede ser utilizado para diversos propósitos. Agresti y Finlay (2009) plantean tres de los más importantes:

- Desvelar modelos de interrelación entre variables.
- Detectar clusters de variables fuertemente correlacionadas y, por lo tanto, redundantes.

- Reducir un número elevado de variables a un número inferior de ellas, siendo estas últimas variables no correlacionadas, lo que en Análisis Factorial se conoce como factores.

El uso del Análisis Factorial, como recurso para determinar evidencias basadas en la estructura interna del test, tiene dos líneas principales (Morales Vallejo, 2011):

1. Análisis Factorial del instrumento, con los ítems que lo forman como unidad de análisis.

2. Análisis Factorial del instrumento junto con otros instrumentos, siendo, en este caso, el test o la escala la unidad de análisis.

Por sí mismo, el Análisis Factorial no prueba la validez del test, pero sí es un buen indicador de si el test mide aquellos aspectos para los que se pensó, aclarando la estructura del instrumento y del constructo (Morales Vallejo, 2011).

Está aceptado que el Análisis Factorial es una herramienta para *explorar* (Harman, 1980) y, como tal, para verificar teorías e hipótesis a partir del análisis de los datos de diferentes experimentos. Esta vertiente de Análisis Factorial se conoce como Análisis Factorial Exploratorio (AFE) y se complementa con una segunda técnica de análisis denominada Análisis Factorial Confirmatorio (AFC). Este último está encuadrado dentro de los modelos de Ecuaciones Estructurales (Batista y Coenders, 2000). Estudia las relaciones entre un conjunto de variables observadas (indicadores) y una o más variables latentes no directamente observables (factores) (Arias, 2008). Para ello, parte de una hipótesis inicial de cómo sería el modelo que define esas relaciones y así, posteriormente, poner a prueba esa hipótesis y contrastarla empíricamente.

Ambas aproximaciones a se han realizado en esta tesis y serán explicadas, sucintamente, a continuación.

Análisis Factorial Exploratorio

El modelo de Análisis Factorial Exploratorio (AFE) expresa los valores esperados de un grupo de n variables observadas como una función lineal de un grupo de m variables latentes, denominadas factores, siendo $m<n$.

Un estudio realizado con AFE da, como resultado, información relevante de cómo se organiza la matriz de correlaciones entre variables, aportando datos sobre aspectos relevantes para un análisis detallado (Morales Vallejo, 2011):

- Número de factores: aporta información sobre el número mínimo de factores que explican la mayor proporción de varianza posible.
- Correlación entre cada variable y cada factor: expresada en la matriz factorial.
- Proporción de varianza explicada por cada factor: calculada a partir de los *autovalores*, esto es el sumatorio del cuadrado de los pesos de cada variable para cada factor.
- Puntuaciones factoriales: estimando las puntuaciones de cada sujeto en cada uno de los factores.

Generalmente, se recomienda que el número de sujetos sea, por un lado, el doble del número de variables y, por otro, que nunca sea inferior a 100 (Kline, 1994). Otros autores (Nunnally, 1978) sugieren, como método de estimación de la muestra necesaria, disponer de una muestra 10 veces mayor que el número de ítems, para asegurar un buen AFE.

Para una muestra suficientemente grande, Amstrong y Soelberg (1968) proponen que ésta sea dividida y que se realicen diferentes análisis factoriales, con las diferentes submuestras obtenidas, para poder verificar de esta forma si la estructura factorial se mantiene estable. En esta línea, Costello y Osborne (2005) recomiendan, para un test de nueva creación, el uso de un AFE inicial en una submuestra del estudio para, posteriormente, realizar un Análisis Factorial Confirmatorio en otra submuestra diferente, verificando así el ajuste de la estructura revelada por el primer análisis.

Si se está interesado en conocer los errores más comunes que se cometen en la práctica del uso del AFE, puede consultarse el artículo de Henson y Roberts (2006).

Para la construcción de nuevos instrumentos de medida, algunos autores recomiendan utilizar el método de Componentes Principales. Sin embargo, otros

proponen el uso de alguno de los métodos de Factores Comunes ya que su estimación de los pesos de las variables es más conservadora (Morales Vallejo, 2011).

Para la determinación del número de factores, el método más seguido es, de nuevo, uno que se encuentra instalado por defecto en uno de los paquetes estadísticos más utilizados, como es el SPSS. El procedimiento Guttman-Kaiser extrae, para su rotación posterior, únicamente los factores que en el primer análisis tienen un autovalor superior a uno. Se trata de una norma muy conservadora y que, especialmente en matrices grandes, tiende a dar un número elevado de factores.

Aunque se han propuesto métodos para determinar el número de dimensiones de una forma menos arbitraria, por ejemplo, eliminando aquellos factores en los que ninguna variable tenga un peso superior a 0,30 (Nunnally, 1978), el método más aceptado es un procedimiento puramente visual, El método Scree Test (Cattel, 1966) (Figura 16).

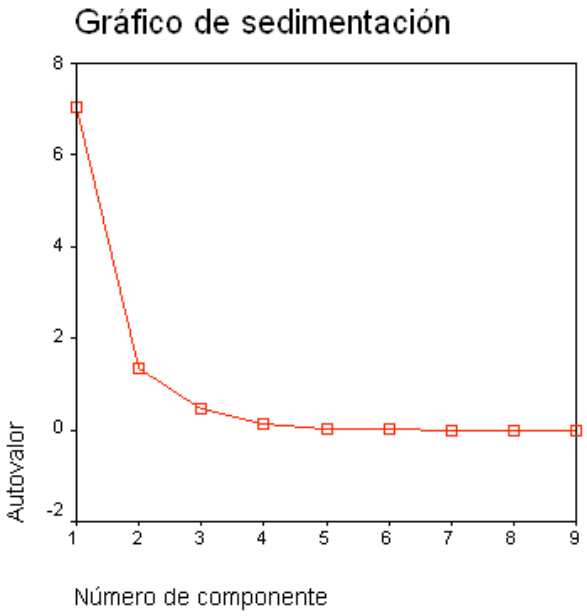


Fig. 16: Ejemplo de Gráfico de Sedimentación

Actualmente, la mejor alternativa es el *Análisis Paralelo* propuesto por Horn (1965). Mediante un procedimiento de simulación Monte Carlo, se genera un número grande de factores y se retienen aquellos factores del estudio real cuyo autovalor supera al 95% de los factores generados aleatoriamente (Morales Vallejo, 2011). Hasta hace poco, la dificultad principal de este análisis radicaba en la inexistencia de un programa adecuado que implementara el procedimiento informáticamente. Ahora mismo existen varios. En concreto, el programa FACTOR,

con el que se realizaron los análisis factoriales del presente estudio, ejecuta análisis en paralelo (Timmerman y Lorenzo-Seva, 2011) y, de hecho, ha sido el método utilizado para la determinación de dimensiones.

De todas formas, a pesar de aplicar estos procedimientos, la interpretación de los factores no es una tarea tan sencilla como podría parecer al principio. Lo habitual no es encontrar una estructura claramente definida. Por el contrario, lo más frecuente es que la primera solución factorial no sea demasiado aclaratoria. En ese caso, se hace necesario un procedimiento que permita rotar la solución factorial para hacerla más interpretable.

Las rotaciones son transformaciones lineales que facilitan la interpretación, sin alterar la proporción de varianza explicada. De hecho, una vez determinado el espacio de los factores comunes, es posible una infinidad de rotaciones de un sistema de coordenadas a otro, sin efecto sobre la adecuación de la solución (Harman, 1980; p. 279). Costello y Osborne (2005) recomiendan el uso del método de rotación oblicua más conocido, *Oblimin Directo*, como mejor método de rotación.

Llegados a este punto, solo queda realizar una interpretación de los resultados obtenidos, teniendo en cuenta el problema del que nos advierte Morales Vallejo (2011; p. 20):

> En la interpretación de los factores, como dimensiones básicas que explican la estructura que subyace a las variables, no hay que olvidar que el análisis factorial, como todos los métodos estadísticos, opera a un nivel puramente estadístico y no conceptual.

Esto implica que, aunque en la mayoría de las ocasiones se describe a los factores como indicadores de una hipotética estructura subyacente, existe una clara indeterminación de la estructura factorial. Conviene no dejar de lado todos los planteamientos que se han venido haciendo históricamente sobre las inferencias causa-efecto realizadas a partir de correlaciones estadísticas (Agresti y Finlay, 2009; McDonald, 1981; Muñiz, 1992).

Un punto de vista acertado a este respecto es considerar el AFE como una forma de sugerir y plantear hipótesis que deben ser contrastadas. Es en este punto donde cobra sentido el uso de la segunda técnica de análisis factorial, el Análisis Factorial Confirmatorio.

Análisis Factorial Confirmatorio

El procedimiento de Análisis Factorial Confirmatorio (AFC) puede ser encuadrado dentro de los Modelos de Ecuaciones Estructurales (Arias, 2008; Batista y Coenders, 2000; Paz, 1996). Estos modelos, nacidos en el marco de la Econometría de los años setenta, fueron posteriormente desarrollados y optimizados por investigadores de la relevancia de Jöreskog (1969) o Bentler (1978).

Una de las potencialidades de este modelo es la posibilidad de introducir el punto de vista confirmatorio en el modelo estadístico. El modelo puede, y debe, ser especificado previamente antes de su estimación. Esa es, en palabras de Arias (2008), una característica esencial del AFC.

Esta especificación inicial del modelo se representa de una manera gráfica mediante lo que se conoce como *Path Diagram* (Duncan, 1975). En la figura 17 puede verse un modelo de representación gráfica mediante un path diagram, tomado de Balboni (2006), a modo de ejemplo por su simplicidad y capacidad descriptiva.

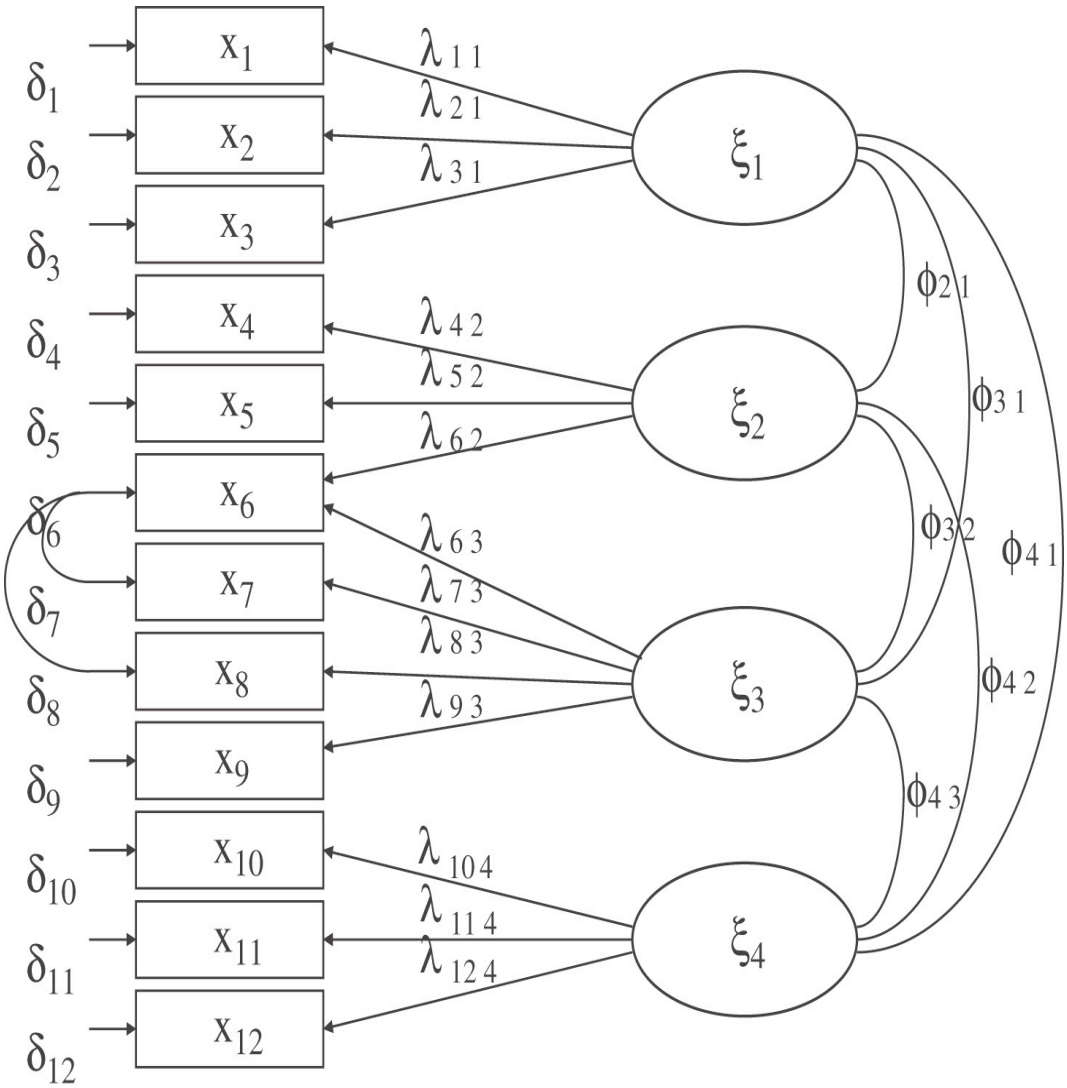


Fig. 17: Ejemplo de Path Diagram de AFC

Arias (2008) expone ordenadamente las condiciones previas que deben darse para poder realizar un AFC con garantías. Estas condiciones se resumen en la tabla 10, tomada del mismo autor.

Condición	Observaciones
Nivel de medida	**Indicadores en nivel de intervalo o de razón (excepcionalmente ordinal)**
Valores por indicador	**Los indicadores deberían tener un mínimo de 4 valores**
Normalidad y outliers	**Distribución normal de los datos, control de los outliers**
Homoscedasticidad	**Corrección mediante normalización o transformaciones**
Datos perdidos	**Tratamiento adecuado de los datos perdidos**
Tipo de relaciones	**Relaciones lineales y aditivas**
Multicolinealidad	**Ausencia de multicolinealidad**
Variables relevantes	**Inclusión dentro del modelo de todas la variables relevantes**
Identificación del modelo	**Modelo supraidentificado**
Número mínimo de observaciones	**Al menos 150 observaciones, o 5 observaciones por cada parámetro a estimar**
Indicadores por variable latente	**Preferible disponer de más de 2 (lo ideal es disponer de al menos 4 o 5)**
Número de indicadores	**El número máximo de indicadores no debería exceder de 20-30**
Varianzas relativas	**Ausencia de matrices** *ill-scaled*

Tabla 8: Condiciones necesarias para realizar un AFC

El autor, en el mismo artículo, propone una descripción más detallada de cada una de estas condiciones subrayando, como queda suficientemente claro a la vista de la tabla de trece condiciones, que la técnica del AFC es de una gran exigencia. El lector interesado en la descripción pormenorizada de estas condiciones puede consultar el texto referido; aquí se expondrán únicamente dos aspectos, dada su especial relevancia.

El primero tiene que ver con la estimación de los parámetros. Cuando se exige un nivel de medida de intervalo o de razón, se está poniendo una condición,

junto con otras, para realizar la estimación de la matriz poblacional Σ, utilizando el método de *Máxima Verosimilitud* (ML). Sin embargo, en el caso de que alguna de las variables indicadoras sea categórica, no está recomendado el uso del procedimiento ML para la estimación de los parámetros. En ese caso, deberá optarse por otros métodos de estimación como *Mínimos cuadrados ponderados* (WLS), *Mínimos cuadrados ponderados diagonalizados* (DWLS), *Mínimos cuadrados ponderados robustos* (WLSMV) o *Mínimos cuadrados no ponderados* (ULS).

El segundo aspecto de relevancia hace referencia a la identificación del modelo. La exigencia de un modelo supraidentificado implica que, en un modelo de AFC, el número de parámetros a estimar ha de ser igual o inferior al número de observaciones (Arias, 2008). Esto significa que el modelo tendrá grados de libertad positivos ($gl \geq 0$).

El indicador más utilizado es χ^2, pero está fuertemente afectado por el tamaño muestral. Por encima de 200 sujetos ($N \geq 200$), el estadístico es casi siempre significativo. Además, con factores ortogonales y correlaciones intrafactor muy semejantes χ^2 no se ve afectado (García Cueto, Gallo, y Miranda, 1998), pero, fuera de esa situación ideal, las diferencias de correlación también afectan al ajuste estimado con este estadístico.

El programa AMOS (Arbuckle, 2006), utilizado para el estudio y análisis factorial confirmatorio de esta investigación, aporta otros indicadores de ajuste. El Índice de ajuste normalizado (NFI) y el Índice de ajuste no normalizado (NNFI), son de los más citados. Ambos funcionan razonablemente bien. Sin embargo, el primero es más sensible a variaciones en el tamaño muestral y al número de parámetros (García Cueto, et al., 1998), por lo que es más recomendable utilizar el segundo. Estos, junto con el Índice de ajuste comparativo (CFI), basado en el tamaño medio de las correlaciones en los datos, deben ser superiores a 0,95 para que se considere el ajuste como bueno. A partir de esa cifra, cuanto más se acerquen estos índices a 1 mejor será el ajuste.

AMOS también aporta los valores del Índice de raíz media del error cuadrático de aproximación (RMSEA) y el de la raíz media del cuadrado de los residuales estandarizados (SRMR). Para ambos, el valor recomendable está por debajo de 0,6. Si su valor es superior a 0,10, se considera que el ajuste es muy pobre.

También se utiliza con frecuencia el criterio de información de Akaike (AIC), aunque para él no hay una medida estandarizada ni hay un criterio de interpretación establecido. En general, se considera que de dos modelos estimados a partir de los mismos datos, aquél con valor de AIC menor es el modelo más recomendable y el de mejor ajuste.

Por último, de una larga serie de índices, de la que se puede tener cumplida información consultando el libro clásico sobre modelos de ecuaciones estructurales editado por Bollen y Long (1993), la salida de resultados de AMOS también aporta el valor del Índice de Hoelter. Este índice indica el tamaño de muestra por debajo del cual χ^2 deja de ser significativo. Se recomiendan valores de al menos 200. Por debajo de 75, el índice señala un ajuste pobre del modelo.

Actualmente, no existe un índice superior al resto. La recomendación más general a este respecto es que se use una combinación de varios como indicador del ajuste.

A pesar de la gran utilidad que muestra el modelo de AFC como herramienta de contraste de hipótesis y confirmación de teorías, su principal limitación como señala Paz (1996) es el hecho de que, a pesar de que un modelo muestre un ajuste excelente, esto no garantiza que la teoría sobre la que se sustenta sea necesariamente válida, ni tampoco que sea la única que se puede formular.

4.2.3 Evidencias basadas en las relaciones con otras variables

Evidencias de validez convergente-discriminante

Cuando se dispone de varios test que miden diferentes rasgos se puede poner en práctica el método de las matrices Multirrasgo-Multimétodo de Campbell y Fiske (1959) para realizar el estudio de la validez de un test en términos de convergencia.

Se habla de *validez convergente* cuando medidas de un mismo rasgo realizadas con diferentes métodos muestran altas correlaciones, por ejemplo, si las puntuaciones de tres test diferentes que miden neuroticismo tienen alta correlación entre si es indicativo de que las medidas de los tres test convergen en el rasgo neuroticismo.

Por otro lado cuando utilizando el mismo método se miden diferentes rasgos, obtener bajas correlaciones es señal de *validez discriminante*. Por ejemplo, se miden, usando el método de autoinforme, tres rasgos diferentes (extraversión, neuroticismo y responsabilidad). Si las correlaciones obtenidas son bajas será indicativo de que, a pesar del método común (autoinforme), no existe influencia de este y los test discriminan perfectamente los diferentes rasgos.

Como señala Paz (1996), la limitación más importante de este método es su esencia heurística y no analítica. No existe ningún criterio de tipo estadístico que permita tomar decisiones sobre la convergencia o divergencia de un test concreto, la decisión queda al albur de la experiencia y conocimiento del investigador que realiza el estudio.

Evidencias de relaciones test-criterio

Esta estrategia de validación es, en cierto modo, la de mejor aplicación práctica ya que, habitualmente, el uso de un test está dirigido a predecir el comportamiento de quien lo responde en alguna variable criterio externa.

De esta forma, a partir del test (predictor) será posible establecer inferencias sobre la actuación del sujeto en una variable externa (criterio).

Esta validez, conocida también como *validez referida a criterio* o *validez criterial*, se operativiza de una forma sencilla mediante el cálculo del coeficiente de correlación lineal producto-momento de Pearson

A este respecto es aconsejable tener en cuenta dos aspectos relevantes para el cálculo de esta correlación. Por un lado, el tamaño muestral recomendado no debe ser inferior a 200 sujetos (N>200), para que las inferencias sean lo suficientemente robustas (García-Cueto, 1993).

Por otro lado, al igual que cualquier otra correlación, el coeficiente de validez así calculado solo indica covariación entre las variables, en este caso test y criterio, pero no necesariamente causalidad.

La cuestión más problemática a la hora de determinar la validez usando este procedimiento radica, normalmente, en la obtención de las puntuaciones del criterio. En primer lugar está la complejidad propia del objeto de medida a la hora de hacer operativo el criterio. La forma en que se mide el criterio tiene una influencia

clave en el estudio de la validez, pudiendo limitar la utilidad del estudio e incluso distorsionando sus resultados (Paz, 1996).

En segundo lugar está el problema de decidir cuándo conviene más obtener esa medida. La situación más habitual que se da, cuando se usa un test como herramienta de selección, es aquella en la que se recogen en primer lugar las medidas de los resultados de todos los aspirantes en el/los test predictores. Posteriormente se eligen a aquellos sujetos con mejores resultados. Pasado un tiempo, en el que han podido mostrar su eficiencia o su ineficiencia en el puesto de trabajo, se realiza la medida del criterio. Este procedimiento se conoce como *validez de pronóstico*.

Además de este enfoque puede darse la necesidad de medir ambos, test y criterio, al mismo tiempo. En este caso se habla de *validez concurrente*.

4.2.4 Resultados

Para el análisis de los datos, en relación con la búsqueda de evidencias de validez de la prueba, se realizaron diferentes pruebas y análisis estadísticos sobre la muestra total descrita en el punto 4.1 (p. 92). Sobre esa muestra se realizaron las codificaciones necesarias en diferentes variables, siempre en función de los requisitos de nivel de medida planteados por las pruebas efectuadas. El procedimiento seguido para realizar estas codificaciones será descrito convenientemente a la hora de plantear los resultados correspondientes.

4.2.4.1 Evidencias de Validez de Contenido

Aquí se ha definido la *Validez de Contenido* como el grado en el que el test constituye una muestra suficientemente representativa de todos los aspectos relevantes del objeto de estudio. Para lograrlo, en la construcción de la herramienta que nos ocupa se optó por la utilización de un panel de expertos que aconsejaran y propusieran los ítems y sus niveles.

Se formó un comité, compuesto por ocho especialistas, dedicado a la determinación de las variables a medir y a la construcción definitiva de los ítems. Cinco de los especialistas eran expertos evaluadores, con experiencia en la conducción de vehículos autocares y también camiones, y con un conocimiento

práctico y suficiente de la aplicación y corrección de la prueba que existía anteriormente, que estaba siendo utilizada hasta ese momento y que fue tomada como punto de partida. Además se incluyó en el panel de expertos la figura de un coordinador que gestionó todo el proceso de convocatoria, organización y resumen de las diferentes reuniones que tuvieron lugar. Por último, dos miembros del Dpto. de Selección de la empresa participaron también en las reuniones del panel de expertos aportando su visión y experiencia desde el conocimiento de diferentes herramientas de evaluación y formación pedagógica.

Tal y como se describe en el punto 3.1.2 (p. 57), donde se mencionan los criterios de elaboración de la prueba, el proceso de construcción de la prueba se realizó en diversas reuniones en las que se fue perfilando la prueba definitiva a partir de dos pilares fundamentales. Por un lado, la prueba antigua que existía hasta ese momento y, por otro, la definición de *Habilidad en la conducción* que se consensuó y aceptó como válida:

> Capacidad de actuar que se desarrolla gracias al aprendizaje , al ejercicio y a la experiencia y que, para el caso de la conducción de vehículos, implica la ejecución de una serie de técnicas y la puesta en marcha de un esquema de conducta relacionado con la Seguridad y la Eficiencia Energética.

Durante las primeras reuniones se consiguió eliminar aquellos aspectos que venían siendo evaluados en la prueba anterior y que se consideró no aportaban elementos de juicio relacionados con la *habilidad en la conducción* así entendida, como se hizo por ejemplo con el ítem *Manejo del tacógrafo* que estaba incluido en la prueba inicial y que fue eliminado en la prueba definitiva.

Tras estas reuniones se llegó al consenso de considerar que los 13 elementos que actualmente conforman la prueba eran aquellos que mejor representaban la habilidad al volante cubriendo conocimientos y habilidades técnicas propias de la conducción de autocares, conductas relacionadas con la seguridad en carretera y habilidades relacionadas con la racionalización del consumo y la conducción energéticamente eficiente.

Una vez determinados los ítems que conforman la prueba se dedicaron las reuniones posteriores, hasta la aprobación de la prueba definitiva, a establecer, claramente y de manera excluyente, los diferentes niveles de la escala para cada uno

de los elementos y determinar aquellas evidencias conductuales que permiten pasar de un nivel a otro en la escala.

Tras la experiencia de cuatro años de uso de la herramienta y las posteriores reuniones, en las que el panel de expertos puso en revisión todos estos aspectos, puede decirse que la prueba, tal y como se aplica actualmente, cubre todos los aspectos posibles relacionados con el constructo *Habilidad en la conducción*, aportando suficiente validez aparente para los candidatos a los que se les aplica esta prueba práctica, pudiendo concluirse que el proceso estructurado de construcción de la prueba ha dado como resultado un test con una aceptable validez de contenido, referido esta a un constructo general que se ha considerado como *Habilidad en la conducción* y que estaría evaluando las competencias, tanto personales como técnicas de un conductor de vehículo industrial, en este caso un autobús.

4.2.4.2 Evidencias de Validez Factorial

Como ya se ha mencionado anteriormente la *Validez de Constructo* persigue obtener evidencias de que el constructo hipotetizado existe y de que además el instrumento elaborado para tal fin es el adecuado para medir dicha variable no observable (Paz, 1996).

En la presente investigación la validez de constructo se ha evaluado utilizando la estimación de la validez factorial a partir de, en primer lugar, un Análisis Factorial Exploratorio (AFE) y, posteriormente, un Análisis Factorial Confirmatorio (AFC) para poner a prueba el ajuste de la estructura obtenida en el AFE previo.

Para ello se utilizó la muestra global descrita en el punto 4.1 (p. 84) formada por 3857 pruebas (N=3857), sin embargo, en el estudio de la validez factorial esta muestra fue segmentada. Costello y Osborne (2005), así como también Henson y Roberts (2006), no recomiendan el uso de AFC para verificar estructuras hipotetizadas previamente cuando se trabaja con test de nueva creación. El uso del AFC supone tener un firme sustrato teórico que respalde la hipótesis formulada. Es por esto que estos autores recomiendan la realización de un análisis exploratorio previo que refuerce el planteamiento de la hipótesis de la estructura. Una vez realizado este AFE la sugerencia de Costello y Osborne es que se realice un análisis confirmatorio para poner a prueba el ajuste de la estructura obtenida, evitando

hacerlo sobre la misma muestra. La recomendación general es que, tras el AFE, el subsiguiente AFC se realice sobre una muestra de la misma población, pero diferente a la muestra original sobre la que se hizo el AFE.

Como consecuencia, al disponer de una muestra suficientemente amplia, para seguir las indicaciones de estos autores y ser lo más estricto posible en la verificación del ajuste de la estructura factorial, se decidió dividir la muestra general en dos submuestras, realizando un primer AFE sobre una de ellas para, posteriormente, comprobar el ajuste de la estructura obtenida realizando un AFC sobre la otra muestra.

En este caso la segmentación de la muestra se llevo a cabo utilizando el método de pares-impares, de tal forma que aquellas pruebas con número de orden par pasaron a formar parte de la *Submuestra 1*, mientras que las pruebas con número de orden impar pasaron a formar parte de la *Submuestra 2*. De esta manera la submuestra 1 quedó conformada por 1929 pruebas ($N_{submuestra1}$=1929) y la segunda por 1928 pruebas ($N_{submuestra2}$=1928) para un total de 3857 pruebas ($N_{muestratotal}$=3857).

Con la submuestra 1 se realizó un AFE utilizando el programa FACTOR (Lorenzo-Seva y Ferrando, 2006). Se determinó un número inicial de 3 factores a extraer, correspondiendo con la estructura inicial que se le dio a la prueba formada por *Conducción Económica, Seguridad* y *Conducción Técnica*. Se utilizó la implementación óptima del Análisis Paralelo (Horn, 1965) como método para determinar el número de dimensiones. El análisis se realizó partiendo de la matriz de correlaciones policóricas tal y como recomiendan Muthen y Kaplan (1985, 1992) que se haga para escalas tipo Likert. La matriz utilizada está representada en la Figura 18. El método de extracción de factores que se aplicó fue Mínimos Cuadrados no Ponderados (ULS) y el método de rotación fue Oblimin Directo normalizado. Se estableció en 100 el número máximo de iteraciones para la convergencia. Por último, se le dio la instrucción al programa de que generara en su salida de resultados las puntuaciones factoriales de los 1929 sujetos en los 3 factores.

STANDARIZED VARIANCE / COVARIANCE MATRIX (POLYCHORIC CORRELATION)
(Polychoric algorithm: Olsson ,1979a, 1979b; Tetrachoric algorithm: Bonett & Price, 2005)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
V 1	1.000												
V 2	0.428	1.000											
V 3	0.338	0.421	1.000										
V 4	0.258	0.230	0.309	1.000									
V 5	0.354	0.277	0.307	0.381	1.000								
V 6	0.357	0.281	0.322	0.383	0.365	1.000							
V 7	0.199	0.150	0.258	0.289	0.184	0.268	1.000						
V 8	0.344	0.281	0.313	0.307	0.285	0.307	0.242	1.000					
V 9	0.184	0.145	0.230	0.277	0.211	0.232	0.240	0.246	1.000				
V 10	0.330	0.309	0.320	0.410	0.418	0.379	0.201	0.314	0.230	1.000			
V 11	0.357	0.361	0.328	0.301	0.357	0.338	0.201	0.336	0.252	0.350	1.000		
V 12	0.528	0.398	0.354	0.324	0.414	0.379	0.221	0.365	0.182	0.369	0.377	1.000	
V 13	0.427	0.361	0.352	0.395	0.447	0.718	0.254	0.377	0.211	0.422	0.412	0.436	1.000

Fig. 18: Matriz de correlaciones policóricas, submuestra 1 (N=1929)

La Figura 19 muestra los valores de adecuación de la matriz de correlaciones. Tanto el resultado del Test de Esfericidad de Bartlett (**p<0,01**), como el valor resultante del test KMO (**0,903),** muestran una buena adecuación de la matriz, indicando de esta forma que la matriz de correlaciones es adecuada para ser utilizada como origen de datos para el AFE.

ADEQUACY OF THE CORRELATION MATRIX

Determinant of the matrix = 0.021199786411926
Bartlett's statistic = 7410.1 (df = 78; P = 0.000010)
Kaiser-Meyer-Olkin (KMO) test = 0.90332 (very good)

Fig. 19: Adecuación de la matriz de correlaciones, submuestra 1 (N=1929)

A continuación se presentan en la Tabla 9 los autovalores obtenidos en la extracción inicial así como la proporción de varianza explicada.

EXPLAINED VARIANCE BASED ON EIGENVALUES

Variable	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
1	4.93510	0.37962	0.37962
2	1.08980	0.08383	0.46345
3	0.95325	0.07333	0.53678
4	0.82175	0.06321	
5	0.75040	0.05772	
6	0.73250	0.05635	
7	0.67999	0.05231	
8	0.64511	0.04962	
9	0.58657	0.04512	
10	0.56408	0.04339	
11	0.51678	0.03975	
12	0.45846	0.03527	
13	0.26623	0.02048	

Tabla. 9: Varianza explicada basada en autovalores, submuestra 1 (N=1929)

Aunque la extracción inicial se forzó a 3 factores puede verse como, utilizando el método de autovalores mayores que 1, se deberían extraer 2. Sin embargo, por el método scree-plot (gráfica no aportada por el programa) el resultado sería de un único factor a extraer[7]. Actualmente la mejor alternativa para determinar el número de factores a extraer es el Análisis Paralelo (Horn, 1965). Ante la dificultad de diferenciar el número de dimensiones por los métodos tradicionales y aprovechando que el Análisis Paralelo viene implementado en FACTOR se decidió tomarlo como referencia para determinar la estructura dimensional. La Figura 20 recoge los resultados del la implementación óptima del Análisis Paralelo (Timmerman y Lorenzo-Seva, 2011) realizada por el programa y donde se señala que la recomendación es extraer un único factor.

```
PARALLEL ANALYSIS (PA) BASED ON MINIMUM RANK FACTOR ANALYSIS
(Timmerman & Lorenzo-Seva, 2011)

Implementation details:

        Correlation matrices analized:                Polychoric correlation matrices
        Number of random correlation matrices:        500
        Method to obtain random correlation matrices: Permutation of the raw data (Buja & Eyuboglu, 1992)


Variable  Real-data      Mean of random   95 percentile of random
          % of variance  % of variance    % of variance

   1       50.6*          16.0             20.2
   2       10.6           14.1             16.7
   3        8.5           12.7             14.9
   4        7.0           11.4             13.1
   5        6.3           10.1             11.6
   6        5.9            8.9             10.3
   7        3.7            7.6              9.2
   8        2.7            6.3              8.1
   9        1.9            5.1              6.9
  10        1.1            3.8              5.8
  11        1.1            2.7              4.9
  12        0.6            1.4              3.3
  13        0.0            0.0              0.0

*  Advised number of dimensions:    1
```

Fig. 20: Análisis Paralelo, submuestra 1 (N=1929)

A continuación se muestra la matriz de pesos factoriales no rotada. (Tabla 10).

[7] Obsérvese que la diferencia en los autovalores de las variables 2 y 3 es muy baja y podría considerarse que la curva se vuelve asintótica después del primer factor.

UNROTATED LOADING MATRIX

Variable	F 1	F 2	F 3	Communality
V 1	0.619	-0.211	-0.246	0.488
V 2	0.544	-0.244	-0.209	0.399
V 3	0.552	-0.173	0.030	0.336
V 4	0.566	0.028	0.323	0.426
V 5	0.586	-0.019	0.051	0.347
V 6	0.670	0.340	-0.014	0.565
V 7	0.380	-0.013	0.228	0.196
V 8	0.532	-0.099	0.071	0.298
V 9	0.370	-0.056	0.289	0.223
V 10	0.591	-0.031	0.120	0.365
V 11	0.575	-0.108	0.007	0.342
V 12	0.650	-0.192	-0.160	0.485
V 13	0.799	0.514	-0.196	0.941

Tabla 10: Matriz factorial no rotada, submuestra 1 (N=1929)

Puede observarse que esta matriz es de fácil interpretación ya que todas las variables saturan en un único factor (F1), con valores por encima de 0,3 y tienen pesos por debajo de esa cifra en los otros dos factores, exceptuando los ítems 6 y 13 que también saturan en el factor 2 (0,340 y 0,514 respectivamente) y el ítem 4 que satura, también por encima de 0,3 en el factor 3 (0,313).

Al forzarse en el análisis la extracción de tres factores el programa ejecutó la rotación de esta matriz factorial utilizando el procedimiento Oblimin Directo para dar una solución factorial rotada. La matriz resultante no muestra una estructura clara ni congruente con el planteamiento inicial según el cual se construyó el test (Tabla. 11).

ROTATED LOADING MATRIX
(loadings lower than absolute 0.300 omitted)

Variable	F 1	F 2	F 3
V 1	0.713		
V 2	0.680		
V 3	0.417		
V 4			0.584
V 5			
V 6		0.678	
V 7			0.411
V 8			
V 9			0.494
V 10			0.342
V 11	0.381		
V 12	0.639		
V 13		1.010	

Tabla 11: Matriz factorial rotada, submuestra 1 (N=1929)

Continuando con la tarea de determinar la dimensionalidad de la prueba y su validez factorial se realizó un análisis factorial de 2º orden utilizando para ello las puntuaciones factoriales de los sujetos obtenidas a partir del AFE inicial realizado con FACTOR.

Para este análisis se utilizó el procedimiento Factor implementado en el paquete estadístico SPSS, utilizando Máxima Verosimilitud como método de extracción. La Tabla 12 recoge la varianza total explicada y la Figura 21 muestra el gráfico de sedimentación para poder realizar el Scree Test.

Varianza total explicada

	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
Factor	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	**2,208**	**73,607**	**73,607**	**1,814**	**60,483**	**60,483**
2	**,421**	**14,048**	**87,655**			
3	**,370**	**12,345**	**100,000**			

Método de extracción: Máxima verosimilitud.

Tabla 12: Varianza total explicada. AF 2º orden, submuestra 1 (N=1929)

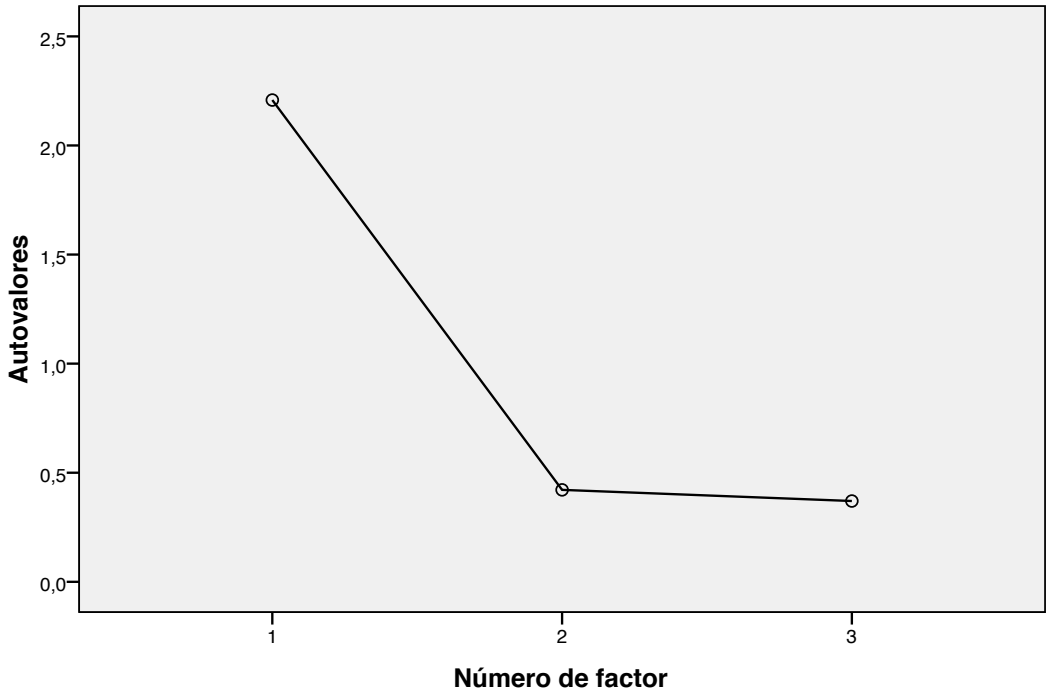


Fig. 21: Gráfico de sedimentación. AF 2º orden, submuestra 1 (N=1929)

En ambos casos se puede comprobar que la estructura en este segundo análisis se ajusta a una estructura unidimensional, de hecho el programa extrae un único factor y por lo tanto ya no ejecuta ningún tipo de rotación de la matriz factorial.

A la luz de estos datos, teniendo en cuenta que además el AFE pone de manifiesto una alta correlación entre los tres factores inicialmente planteados (Tabla 13), parece sensato pensar en la unidimensionalidad de la prueba y plantear un modelo de un único factor en el que la prueba estaría midiendo *Habilidad en la conducción*. Partiendo de esa idea el AF Confirmatorio se modela para verificar el ajuste de los datos de la submuestra 2 a una estructura unidimensional.

INTER-FACTORS CORRELATION MATRIX

Factor	F 1	F 2	F 3
F 1	1.000		
F 2	0.624	1.000	
F 3	0.608	0.580	1.000

Tabla. 13: Matriz de correlación entre factores, submuestra 1 (N=1929)

Para la realización del Análisis Factorial Confirmatorio se utilizó el programa AMOS en su versión 7.0 (Arbuckle, 2006). Se tomó como muestra la submuestra número 2 (N=1928) obtenida a partir de la división de la muestra original. En la Figura 22 puede verse el Path Diagram utilizado, en el que se modela un único factor con pesos en las trece variables y en el que se han tenido en cuenta las intercorrelaciones de los errores de medida para mejorar el ajuste.

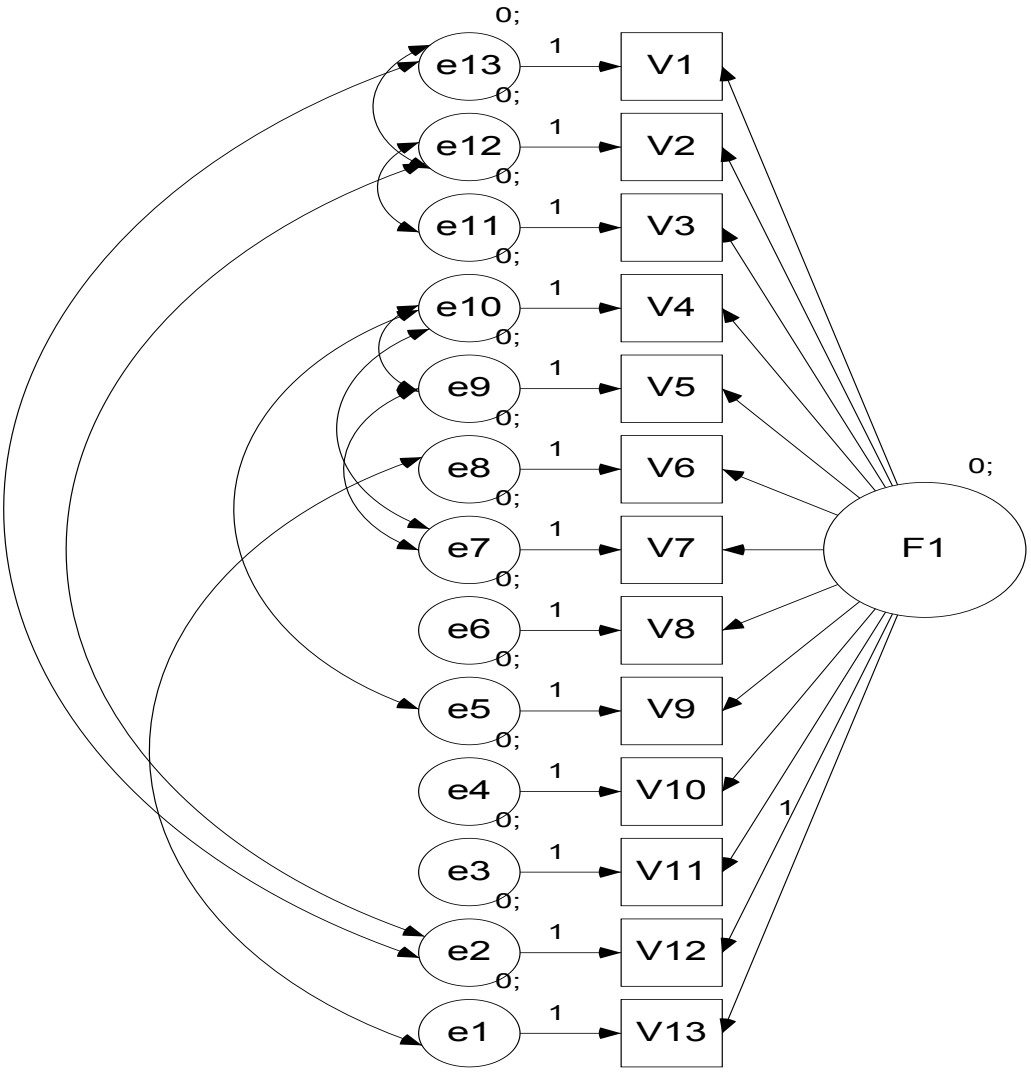


Fig. 22: Path Diagram para AFC, submuestra 2 (N=1928)

Tras la ejecución del análisis se obtuvo el sumario de todos los indicadores de ajuste del modelo que proporciona el programa AMOS. Aquí se presentan aquellos que se consideran más significativos y robustos (Tabla 14). Tal y como se comentó anteriormente no existe un único indicador que destaque sobre el resto y que pueda considerarse como único indicador fiable, por lo que suele acudirse a la combinación de varios de ellos para determinar la bondad del ajuste. El listado de todos los indicadores obtenidos está disponible en el anexo de esta memoria de tesis.

Indicador de Ajuste	Valor obtenido
χ^2 CMIN	**p<0,000**
χ^2 /DF CMIN/DF	**5,667**
Índice de Ajuste Normado NFI	**0,967**
Índice de Ajuste Comparado CFI	**0,972**
Indice Incremental de Ajuste IFI	**0,972**
Error Medio de Aproximación Cuadrática MRSEA	**0,049**
Hoelter .05	**453**
Hoelter .01	**508**

Tabla 14: Indicadores de ajuste del modelo de AFC, submuestra 2 (N=1928)

Los dos primeros indicadores que se presentan son χ^2 y la razón de el estadístico chi cuadrado y los grados de libertad χ^2/DF. El primero se denomina CMIN y, por extensión, el segundo CMIN/DF. Para CMIN lo recomendable para poder expresar un buen ajuste es que su valor no sea significativo (p>0,05). Para CMIN/DF la indicación de un buen ajuste se sitúa en valores por debajo de 5.

En la tabla se muestra también el Índice de Ajuste Normado (NFI), el Índice de Ajuste Comparado (CFI) y el Índice Incremental de Ajuste (IFI). Los tres son modelos de ajuste comparativo para una sola muestra y los tres indican que existe un buen ajuste del modelos cuando toman valores por encima de 0,95.

Además se presenta un indicador basado en el parámetro de no centralidad. Se considera que el Error Medio de Aproximación Cuadrática (MRSEA) de Steiger y Lind revela un buen ajuste cuando su valor es inferior a 0,06 (RMSEA<0,06).

Por último, se presenta el Índice Hoelter. Como se explicó previamente, su valor indica el tamaño de muestra a partir del cual χ^2 empieza a ser siempre significativo. Expresa por tanto un límite superior, por debajo de los tamaños muestrales resultado de este índice la información de χ^2 es fiable. La recomendación general es que su valor sea superior a 200 y que solo se haga constar su resultado en el caso de que χ^2 arroje un resultado significativo, como sucede en esta

investigación. Aquí se presenta su valor tanto para niveles de significación alpha=0,05 como para alpha=0,01.

Como se puede comprobar en la tabla, todos los indicadores están en valores asociados a un buen ajuste salvo los indicadores de χ^2. Como es conocido, χ^2 es una medida de ajuste que está muy influida por el tamaño muestral. En análisis con tamaños muestrales por encima de los 200 sujetos χ^2 es prácticamente siempre significativo. Para eso se presenta el Índice Hoelter, en el que se puede comprobar que para confiar en la medida del ajuste realizada con χ^2 la muestra debería ser de tamaño inferior a 453 sujetos, número que se supera con creces, por lo que el resultado diferencial de χ^2, con respecto al resto de indicadores, no es extraño.

En definitiva, tras el exhaustivo estudio de la validez de constructo realizado, todos los datos obtenidos permiten sostener la hipótesis de que la prueba práctica de evaluación de conductores es un test que evalúa una única dimensión que podría ser denominada como *Habilidad en la conducción*.

4.2.4.3 Otras evidencias de validez

Para esta investigación no se realizó un análisis exhaustivo y en profundidad de la Validez Predictiva. Se pudieron recabar datos de siniestralidad de los conductores de la Organización ALSA, pero los datos provistos no estaban codificados de una manera clara como para ser usados directamente, no se habían obtenido específicamente para realizar una investigación, por lo que carecían de los controles necesarios y no eran fácilmente interpretable lo que también dificultaba su manejo.

Tampoco se disponía de otra información relevante como podrían ser el tipo de servicio realizado, el tipo de siniestro, la determinación de la culpa, la localización del mismo, el tipo de vehículo que se conducía o una descripción de la jornada laboral del conductor en el momento del percance. En este caso únicamente se disponía del número total de siniestros de un conductor en un periodo de tiempo. En el caso general este periodo de tiempo correspondía a un año.

Se consideró que los datos, obtenidos de esta forma, no eran lo suficientemente informativos y no aseguraban los controles mínimos como para realizar un estudio de validez predictiva exhaustivo, que permita determinar la capacidad de la prueba para detectar candidatos con unas habilidades en la

conducción relacionadas con una mayor probabilidad de tener un accidente, a partir de ellos.

Queda, por tanto, pendiente el establecer los criterios y controles necesarios para recabar la información suficiente sobre siniestros de tal manera que permita realizar un estudio sistemático y profundo de la capacidad predictiva del test.

Los que si se realizó fue un estudio de Regresión Logística, complementario a la validez, para tratar de determinar la *Sensibilidad* de la prueba.

Sensibilidad de la prueba

Como estudio de la sensibilidad de la prueba y de su capacidad de clasificación de los candidatos se tomaron, de la muestra global (N=4425), aquellas pruebas de las que se tenían datos de accidentalidad, durante el año 2010, del candidato que las había realizado.

Los datos provienen de la tabla de siniestros proporcionada por la empresa ALSA y que se maneja habitualmente para el control y formación de sus trabajadores. En dicha tabla únicamente constaba el DNI del trabajador y su número de siniestros en el periodo de un año. El dato del DNI fue utilizado únicamente a efectos de poder relacionar los datos de siniestralidad con el sujeto que había realizado una prueba en concreto. En todo momento se preservó la confidencialidad de la información y se respetó escrupulosamente la Ley de Protección de Datos.

Tras cruzar la tabla de siniestros con la tabla de pruebas realizadas se obtuvieron 390 pruebas de las que se disponía del dato de siniestralidad de su conductor[8].

Con estas 390 pruebas se realizó un estudio utilizando el procedimiento de Regresión Logística, implementado en SPSS, tras haber codificado previamente la variable dependiente *Siniestralidad* otorgando el valor 0 a los sujetos con 2 o menos siniestros y el valor 1 a los sujetos con más de 2 accidentes. La primera categoría se definió como *Baja siniestralidad* y la segunda como *Alta siniestralidad*.

[8] Obviamente, para disponer del dato de siniestralidad era necesario que el conductor fuera trabajador de la empresa, los que implica que su prueba de evaluación debe ser apta (condición necesaria para acceder a este puesto de trabajo), por lo que la media de las puntuaciones de esta muestra de pruebas (M=24,77) es más alta que la media de la muestra general.

A continuación se muestra la tabla de clasificación obtenida directamente de los resultados obtenidos por los 390 candidatos. Con ella se inicia el análisis de regresión logística utilizando el método *Introducir*, con 20 iteraciones para la convergencia y 0,5 como valor de corte.

Tabla de clasificación [a,b]

			Pronosticado		
			Nº de Siniestros		
	Observado		Baja siniestralidad	Alta siniestralidad	Porcentaje correcto
Paso 0	Nº de Siniestros	Baja siniestralidad	**234**	**0**	**100,0**
		Alta siniestralidad	**156**	**0**	**,0**
	Porcentaje global				**60,0**

a. En el modelo se incluye una constante.

b. El valor de corte es ,500

Tabla 15: Tabla de clasificación observada (N=390)

Tras realizarse la regresión introduciendo las tres escalas iniciales de la prueba (Conducción Económica, Seguridad y Conducción Técnica) como variables independientes se obtiene la siguiente tabla de clasificación pronosticada (Tabla 16).

Tabla de clasificación [a]

			Pronosticado		
			Nº de Siniestros		
	Observado		Baja siniestralidad	Alta siniestralidad	Porcentaje correcto
Paso 1	Nº de Siniestros	Baja siniestralidad	**227**	**7**	**97,0**
		Alta siniestralidad	**145**	**11**	**7,1**
	Porcentaje global				**61,0**

a. El valor de corte es ,500

Tabla 16: Tabla de clasificación pronosticada (N=390)

En ella puede comprobarse como la prueba de evaluación clasifica correctamente al 97% de los sujetos de *Baja siniestralidad*, sin embargo, desgraciadamente, no tiene tanta sensibilidad para detectar y clasificar a los sujetos de *Alta siniestralidad* y únicamente clasifica correctamente a un 7% de ellos.

Por lo tanto, puede concluirse que la prueba es sensible y clasifica bien en niveles bajos de siniestralidad, pero falla a la hora de clasificar a los sujetos de alta siniestralidad.

4.3 Análisis de Fiabilidad

4.3.1 Modelo Lineal Clásico

Toda medida ya sea psicológica, física o biológica lleva asociada un cierto error de medida. Con esta frase el catedrático de Psicometría de la Universidad de Oviedo, el profesor José Muñiz, refleja lo expuesto por Robert Thorndike en su clásico trabajo sobre fiabilidad (Thorndike, 1951).

> Whenever we measure anything, whether in the physical, the biological, or the social sciences, that measurement contains a certain amount of chance error. The amount of chance error may be large or it may be small, but it is universally present to some extent.

Por extensión, cuando en Psicología se utiliza algún tipo de herramienta, ya sea un test o una escala, para medir una variable concreta esa medida tampoco está libre de error. La fiabilidad de un test viene entonces determinada por su capacidad de no solo medir cometiendo un nivel aceptable de error sino también de hacerlo de manera consistente en el tiempo. Esto implica que un instrumento de medida debe ser preciso (debe medir con la menor cantidad de error posible), pero también ha de ser fiable (ha de poder reportar resultados similares en medidas consecutivas a lo largo del tiempo de una variable que se considera estable). En todo caso, esto no debe confundirse con la estabilidad de la variable medida, para un análisis detallado de este tema Muñiz (Muñiz, 1992) recomienda la revisión del trabajo de Silva (Silva, 1989). Es evidente que si un test reporta resultados diferentes en distintas medidas de una variable que cambia con el tiempo, que no es estable, esta variabilidad no puede ser achacada a una falta de fiabilidad del test.

En definitiva, el término fiabilidad engloba una serie de procedimientos utilizados para estimar el error que se comete al medir (Muñiz, 1996).

Antes de pasar a definir la medida utilizada aquí para estimar la fiabilidad conviene añadir un detalle. En la Teoría Clásica de los Test el coeficiente de fiabilidad está afectado por la longitud del test y por la variabilidad de la muestra.

En el primer caso, un aumento del número de ítems supone un cambio en sus varianzas (empírica, verdadera y de error). En la deducción de la fórmula de Spearman-Brown para el cálculo del coeficiente de fiabilidad cuando aumenta la longitud de un test se puede comprobar que la varianza verdadera (numerador) aumenta proporcionalmente más que la varianza empírica (denominador) al aumentar la longitud del test. Esto es lo que hace que la fiabilidad aumente al aumentar el número de ítems. Al principio de una manera más acusada y de una manera más efectiva en test con un número reducido de ítems.

Para el segundo caso, el coeficiente de fiabilidad aumenta cuando aumenta la variabilidad de la muestra. Recuérdese que el coeficiente de fiabilidad es la expresión de la correlación entre dos formas paralelas de un test y es conocido que la correlación está afectada directamente por un aumento de la variabilidad.

Esto tiene una implicación muy importante para el modelo lineal clásico, ya que de lo expuesto anteriormente puede deducirse que un test no tiene un único coeficiente de fiabilidad, este siempre va a depender de la muestra en la que se aplique. Por lo tanto, esto resta poder de generalización a los resultados obtenidos ya que el dato sobre la fiabilidad de un test solo será generalizable a la población de la que haya sido extraída la muestra.

4.3.2 Coeficiente Alpha (α)

Sin ningún lugar a dudas el coeficiente α, propuesto por Cronbach (1951), es uno de los índices más comúnmente utilizados en investigación para determinar la fiabilidad de un instrumento de medida, sea cual sea el campo de trabajo (Ballester, Gil, Gómez, y Gil, 2010; Nogueira, Godoy, Gavino, y Romero, 2009; Recio, Cuadrado, y Ramos, 2007; Torres y Cuesta, 1992).

Alpha es un indicador de la consistencia interna del test ya que refleja el grado en que correlacionan los ítems del test . Matemáticamente se expresa por:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma\sigma_j^2}{\sigma_x^2}\right)$$

Donde *n* es el número de ítems, σ_j^2 es la varianza del ítem j y σ_x^2 es la varianza de las puntuaciones totales del test. Puede encontrarse una excelente deducción de esta fórmula en García Cueto (1993)

Existen otros métodos de estimación de la fiabilidad basados en la consistencia interna, (Flanagan, 1937; Guttman, 1945; Kuder y Richardson, 1937; Rulon, 1939), pero todos ellos son, en realidad, casos particulares de Alpha, toda vez que puede demostrarse que el coeficiente α es el valor medio de todos los coeficientes de fiabilidad calculados para todas las mitades posibles del test.

Conviene por último indicar que Alpha no debe se utilizado como indicador único de la unidimensionalidad de un test. Este punto se tratará con mayor detalle más adelante, baste señalar aquí que el coeficiente α equivale al coeficiente de fiabilidad únicamente en el caso de que los ítems del test sean paralelos y estén midiendo la misma dimensión. En cualquier otro caso Alpha es un estimador por defecto del coeficiente de fiabilidad y su valor siempre será menor.

En todo caso esto no implica que valores altos de consistencia interna presupongan unidimensionalidad del test. Si el test es unidimensional es esperable que la consistencia interna sea alta, pero el caso contrario no siempre se cumple. Se verá en puntos posteriores cómo, para determinar la unidimensionalidad de la prueba, serán necesarias otras pruebas además de la constatación de la consistencia interna del test.

4.3.3 Resultados desde la Teoría Clásica de los Test

La estimación del a fiabilidad del test se realizó utilizando Alpha como medida de la consistencia interna de la prueba.

Se calculó Alpha utilizando el programa FACTOR (Lorenzo-Seva y Ferrando, 2006). Se trata de una aplicación informática que realiza Análisis Factorial Exploratorio (AFE) en un entorno relativamente sencillo de uso y aportando procedimientos estadísticamente más avanzados que aquellos que se presentan tradicionalmente en los paquetes estadísticos comerciales.

El AFE proporciona, por si mismo, indicadores de la consistencia interna de un test muy semejantes a Alpha e, incluso, ya de por si constituye un indicador de la consistencia del test a partir del estudio de la matriz de correlaciones, la estimación del número de factores y el porcentaje de varianza explicada por ellos (Muñiz, 1992).

En FACTOR el cálculo de la fiabilidad y consistencia interna se realiza a partir del cuadrado del error estándar de las puntuaciones factoriales rotadas (Mislevy y Bock, 1990) lo que aporta un valor más ajustado de Alpha.

En este caso, el análisis de la escala completa arroja un α=0,899, ligeramente superior al calculado mediante el procedimiento Reliability de SPSS (Tabla 17)

Factor	Varianza	Alpha
V 1	**5,046**	**0,899**

Tabla 17: Resultados de α en FACTOR, escala completa

Además, se ha calculado el valor de Alpha en las tres escalas en las que, por motivos operativos, se dividió la prueba en el momento de su construcción y que ya se mencionaron en el capítulo 3 (Conducción Económica; Seguridad y Conducción Técnica).

La Tabla 18 muestra la comparación de los resultados de Alpha, para cada escala, calculados por ambos procedimientos.

Escala	α FACTOR
Conducción Económica	**0,761**
Seguridad	**0,693**
Conducción Técnica	**0,780**
Total	**0,899**

Tabla 18: Valor de α en las tres escalas de la prueba de evaluación

Se puede observar que los valores de consistencia interna de la prueba, en general, son elevados, siendo ligeramente superiores cuando el cálculo se realiza utilizando el procedimiento establecido en FACTOR, a partir del Análisis Factorial Exploratorio.

4.4 Análisis de los ítems

4.4.1 Fundamentos del Análisis de los Ítems

El profesor García Cueto (1993) identifica y define los pasos necesarios para una correcta construcción de un test. Los pasos iniciales suponen un análisis exhaustivo del rasgo, es decir aclarar la naturaleza y características del objeto a medir, y la elaboración de un conjunto suficiente de ítems para su uso y estudio.

Necesariamente ese conjunto de ítems debe pasar por una tercera fase que implica su aplicación, estudio y selección con el objetivo de elegir aquellos que son mejores y más adecuados para la medida del rasgo. El análisis de los ítems consiste, entonces, en estudiar aquellos de sus parámetros cuyas características estén relacionadas con las propiedades y finalidad del test (Muñiz, Fidalgo, García-Cueto, Martínez, y Moreno, 2005).

Para realizar este análisis se le indican, al constructor de test, dos pasos que debe cumplimentar. Por un lado, un primer análisis de tipo *subjetivo* que, normalmente, será realizado por expertos conocedores de los principales aspectos relacionados con el rasgo a medir. En este paso aquellos ítems que no están redactados con claridad, aquellos que por su ambigüedad pueden ser malinterpretados o los que no muestran ninguna relación con el campo evaluado, son eliminados o corregidos para eliminar sus posibles errores de construcción.

En segundo lugar deberá realizarse un análisis *objetivo* en el que se estudiarán matemáticamente aquellas características de los ítems que tengan mayor influencia directa en la mejora de las propiedades del test como instrumento de medida (Muñiz, 1992; Muñiz, et al., 2005).

Toda vez que diversos autores han destacado la dependencia del análisis de los ítems del modelo, TCT o TRI, sobre el cual se haya basado la construcción del test (Bechger, Maris, Verstralen, y Béguin, 2003; Ellis y Mead, 2008; Muñiz, et al., 2005; Sinar y Zickar, 2002), ambas perspectivas serán analizadas aquí.

4.4.2 Perspectiva clásica del Análisis de Ítems

Para empezar, desde el modelo clásico se considera que los indicadores básicos a estudiar deben ser:

- Dificultad.
- Discriminación.
- Análisis de distractores.
- Fiabilidad.
- Validez.
- Dimensionalidad.

Dificultad de los ítems

Básicamente la dificultad de un ítem viene determinada por la proporción de personas que lo aciertan del total de los que lo intentaron responder. Se entiende fácilmente que el número de aciertos, por si mismo, no es indicador de la dificultad y que es necesario poner en relación ese dato con el total de respuestas.

Cabe destacar el hecho, a la luz de la fórmula de cálculo, de que cuanto mayor sea el numerador (número de aciertos) mayor será el valor del índice. Esto implica que el índice de dificultad muestra una relación inversa, cuanto mayor es el valor del índice menor es la dificultad del ítem. Se podría definir más bien como un índice de facilidad más que de dificultad.

De nuevo hace aparición la limitación de la dependencia muestral en el modelo clásico. El índice de dificultad así calculado no constituye una propiedad intrínseca del ítem (Muñiz, 1992), sino que depende de la capacidad de los sujetos de la muestra a la que se aplica el ítem para calcular su dificultad. Un mismo ítem

puede resultar muy difícil para aquellos sujetos con una competencia baja y a su vez ser muy fácil para aquellos que muestran un nivel más alto de competencia.

Discriminación de los ítems

Se considera que un ítem tiene un alto poder discriminativo si permite diferenciar entre aquellos sujetos que puntúan alto en el test y aquellos que puntúan bajo.

En otras palabras, si un ítem discrimina bien será acertado por aquellos que muestran una buena ejecución en el test y, por el contrario, será fallado por los que no son tan buenos a la hora de responder al test completo.

Así definida la discriminación de un ítem supone que debe de existir una correlación entre las puntuaciones en el ítem y las puntuaciones de los sujetos en el test.

Actualmente, lo más habitual para la estimación de la discriminación de un ítem es el uso de la correlación ítem-test. El *índice de discriminación* así calculado es la expresión de la homogeneidad con la que el ítem mide con respecto al test, es decir, hasta que punto el ítem está midiendo la misma variable que el test. Esta medida, así calculada, obliga a tener en cuenta dos aspectos principales.

En primer lugar, el hecho de que sea una medida de la homogeneidad puede llevar a interpretaciones ambiguas del índice como indicador de la discriminación del ítem.

En segundo lugar, en el cálculo de la correlación será necesario eliminar la influencia del ítem sobre el valor total. Nótese que el ítem forma parte del test total, al calcular la correlación entre ambos se está correlacionando el test con una parte de si mismo. Por tanto, para no sobreestimar el valor de esa correlación debe aplicarse una corrección tal que elimine el efecto del ítem estudiado sobre la correlación total ítem-test. Para ello en la literatura sobre este particular se propone la siguiente fórmula:

$$\varrho_{j(x-j)} = \frac{\varrho_{jx}\sigma_x - \sigma_j}{\sqrt{\sigma_j^2 + \sigma_x^2 - 2\varrho_{jx}\sigma_j\sigma_x}}$$

Donde el subíndice *j* se refiere al ítem analizado y el subíndice *x* al test total.

Por último, habrá que tener muy en cuenta el tipo de escala sobre la que está construido el test. Este es un dato clave para determinar cuál ha de ser el método de cálculo de la correlación más adecuado ya que, como es conocido, la correlación producto-momento de Pearson exige, para su cálculo, variables continuas y distribuciones normales bivariadas, por lo tanto no es el método de cálculo de la correlación adecuado para todo tipo de escalas.

Para facilitar la elección correcta se propone el siguiente cuadro (Muñiz, et al., 2005):

Ítem	Test		
Nivel de medida	Dicotómico	Dicotomizado	Continuo
Dicotómico	Correlación Φ	Correlación Φ	Correlación Biserial-Puntual
Dicotomizado		Correlación Tetracórica	Correlación Biserial
Continuo			Correlación Pearson

Tabla 19: Correlación recomendada en función del nivel de medida.
Tomado de Muñiz et al. (2005)

Análisis de los distractores

El análisis de los distractores de un test consiste en verificar que aquellas alternativas falsas que contengan los ítems, es decir las que no se corresponden con la alternativa correcta, son igual de *atractivas* para aquellas personas que desconocen la respuesta correcta al ítem.

Es decir, se trata de verificar que no existan alternativas tan poco verosímiles que, aún no conociendo la respuesta al ítem, nunca son respondidas por ser consideradas claramente falsas.

Su cálculo, a partir de la proporción de elecciones de cada alternativa verificada con una prueba estadística de X^2, tiene sentido para test en los que una alternativa es correcta frente a las otras, es decir, en un análisis pertinente en test de acierto-fallo. Este no es el caso de la prueba de evaluación que nos ocupa.

Fiabilidad de los ítems

La fiabilidad de cada ítem contribuye a la fiabilidad del test completo y puede, además, utilizarse como herramienta de selección de ítems en el proceso de construcción de una prueba.

El cálculo del *índice de fiabilidad* se realiza a partir de:

$$IF = S_i ID_i$$

Donde, S_i es la desviación típica de las puntuaciones del ítem i e ID_i es el índice de discriminación del ítem i.

Por último, para finalizar el análisis de ítems desde la perspectiva de la Teoría Clásica de los Test, se señalará que tanto la Validez como la Unidimensionalidad, serán estudiadas con detalle más adelante.

4.4.2 Resultados

El análisis de ítems se realizó sobre la misma muestra que fue utilizada para el análisis de fiabilidad, con N=3857. En este caso se realizó el análisis utilizando el procedimiento RELIABILITY implementado en SPSS 17, pidiendo el cálculo de los estadísticos descriptivos para todos los elementos de la escala.

Dificultad

Como ya se describió, el índice de dificultad es un indicador de lo fácil o difícil que resulta acertar un ítem concreto. Es por esto que no es un indicador que deba ser utilizado sobre escalas de actitudes o test de personalidad, ya que, por definición, no están construidos en términos de acierto-fallo. Es complejo estimar la dificultad de un ítem que está formulado en un continuo sin que exista la posibilidad de respuestas correctas o incorrectas.

En el caso de ítems politómicos, como escalas tipo Likert similares a la utilizada para esta prueba, el índice de dificultad puede estimarse tomando la media de las puntuaciones en todas las categorías del elemento (López Pina, 2005). Este dato será indicador, más que de la dificultad del ítem, de la probabilidad de dar una respuesta determinada en función del nivel aptitudinal o del grado actitudinal (Muñiz, et al., 2005).

En la Tabla 20 se presentan los estadísticos de los trece elementos que componen la prueba. Recuérdese que todos los ítems están en una escala de 4 niveles en un intervalo de 0 a 3 puntos en el que 0 indica el nivel más bajo en la ejecución y 3 el nivel más alto.

Estadísticos de los elementos

	Media	Desviación típica	N
EcoRevoluciones	**1,69**	**,854**	**3857**
EcoRetarder	**1,27**	**,948**	**3857**
EcoFreno	**1,75**	**,586**	**3857**
EcoAnticipacion	**1,71**	**,678**	**3857**
SegControl	**1,83**	**,697**	**3857**
SegRetrovisor	**1,77**	**,638**	**3857**
SegDistancia	**1,91**	**,573**	**3857**
SegVolante	**1,88**	**,766**	**3857**
SegSenalizacion	**1,90**	**,718**	**3857**
TecFluidez	**1,81**	**,675**	**3857**
TecMandos	**1,43**	**,820**	**3857**
TecCambio	**1,66**	**,839**	**3857**
TecManiobra	**1,65**	**,885**	**3857**

Tabla 20: Estadísticos descriptivos de los ítems

Se puede observar que los valores medios de los trece elementos están por debajo de 2 y son bastante homogéneos. Únicamente destaca el valor medio del ítem número 2 EcoRetarder, el cual con M=1,27 resulta ser el ítem más difícil.

Discriminación

Como se indicó en secciones anteriores, el poder discriminativo de un elemento se determina por la correlación entre las puntuaciones obtenidas en el ítem y las puntuaciones totales del test. Esta correlación ítem-test debe ser corregida para eliminar el efecto del elemento sobre la puntuación total, que elevaría de manera espuria el valor de la correlación (García-Cueto, 1993; Muñiz, 1992).

La Tabla 21 muestra los estadísticos total-elemento en los que se recoge el valor de la correlación ítem-test corregida en la columna señalada en rojo en la tabla.

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
EcoRevoluciones	20,56	33,540	,620	,875
EcoRetarder	20,97	33,393	,558	,880
EcoFreno	20,50	35,707	,622	,876
EcoAnticipacion	20,54	35,189	,592	,877
SegControl	20,42	34,879	,613	,876
SegRetrovisor	20,48	35,229	,631	,875
SegDistancia	20,34	37,521	,365	,886
SegVolante	20,36	34,879	,546	,879
SegSenalizacion	20,35	36,661	,372	,887
TecFluidez	20,43	34,947	,627	,875
TecMandos	20,82	33,922	,609	,876
TecCambio	20,59	33,275	,665	,872
TecManiobra	20,60	32,734	,681	,871

Tabla 21: Estadísticos total-elemento

En general los valores son aceptables. Los diferentes elementos de la escala muestran valores altos de correlación con el test total. Recuérdese que este índice, así calculado, es también un indicador de la homogeneidad de medida del test, sin que por ello sea muestra de la existencia de unidimensionalidad. Solamente los ítems 7 (SegDistancia) y 9 (SegSeñalización) muestran valores de discriminación sensiblemente más bajos que el resto (r_{7t}=0,365 y r_{9t}=0,372), mientras que los demás componentes de la escala están en valores de correlación ítem-test por encima de 0,5, siendo el ítem número 13 (TecManiobra) el de mayor poder discriminativo (r_{13t}=0,681).

Fiabilidad

La fiabilidad individual de cada ítem contribuye a la fiabilidad del test completo.

A partir de los datos obtenidos en el estudio de las propiedades de los ítems se ha calculado la siguiente tabla (Tabla 22) donde se recogen los valores de fiabilidad para cada uno de los trece elementos de la escala total.

Elemento	Fiabilida	Elemento	Fiabilidad	Elemento	Fiabilidad
EcoRevoluciones	**0,529**	SegRetrovisor	**0,402**	TecMandos	**0,499**
EcoRetarder	**0,529**	SegDistancia	**0,209**	TecCambio	**0,558**
EcoFreno	**0,364**	SegVolante	**0,418**	TecManiobra	**0,603**
EcoAnticipación	**0,401**	SegSeñalización	**0,267**		
SegControl	**0,427**	TecFluidez	**0,423**		

Tabla 22: índices de fiabilidad de los elementos

El cálculo se ha realizado utilizando:

$$IF = S_i ID_i$$

Donde la fiabilidad del ítem viene dada por el producto entre S_i que es la desviación típica de las puntuaciones del ítem i e ID_i que es el índice de discriminación del ítem i.

4.5 Teoría de Respuesta a los Ítems

4.5.1 Principios Generales

Hasta este punto se ha podido poner de manifiesto en las secciones anteriores de este capítulo que, desde el enfoque clásico de la medición psicológica, existe una profunda relación entre las propiedades psicométricas del instrumento de medida y la muestra utilizada para determinarlas.

Esto, como ya se ha comentado, afecta directamente al coeficiente de fiabilidad del test, pero no solo sino también a otros valores como pueden ser el índice de discriminación o el índice de dificultad.

Hambleton señala esta dependencia del grupo normativo como el primero de los defectos, (shortcomings, en la versión original de su artículo), achacables a los procedimientos clásicos. Resume este inconveniente de la siguiente manera:

> The average level of ability and the variability of ability scores in a examinee group influence the value of the item statistics, and reliability and validity statistics too, often substantially (Hambleton, 1990)

Esta limitación implica, como ya se señaló en otra ocasión, que no solo existen tantos índices y coeficientes como aplicaciones diferentes se hagan del test, sino que los resultados son difícilmente generalizables cuando se aplica a una muestra de sujetos con características diferentes a las de aquella muestra sobre la que se determinaron las propiedades psicométricas del test.

Además de esta traba inicial el enfoque clásico presenta un segundo inconveniente a la hora de establecer comparaciones. Desde este enfoque es muy

difícil poder comparar las medidas, en una determinada aptitud, de sujetos que han sido evaluados en la misma variable pero con diferentes instrumentos, algo así como si los pesos registrados para dos individuos en dos básculas diferentes no pudieran ser comparables. Inicialmente la única posibilidad de comparación exigiría haber aplicado dos test estrictamente paralelos, caso que, conforme a los requisitos conocidos, es extremadamente complejo y caro de conseguir.

Por último, el tercer defecto que el profesor Hambleton señala en la Teoría Clásica de los Test es el hecho de que no proporciona ninguna clave para poder determinar o predecir el resultado de enfrentar a un sujeto con un ítem concreto.

Así las cosas, la Teoría de Respuesta a los Ítems (TRI) constituye un nuevo enfoque que permite resolver ciertos problemas de la medición psicológica, ya mencionados, inabordables desde la perspectiva clásica (Muñiz, 1990).

Cabe señalar que la TRI no viene a eliminar los presupuestos ni el desarrollo de la perspectiva clásica, sino que complementa añadiendo nuevos puntos de vista y asunciones que permiten dar respuesta a cuestiones que quedaban sin explicar en la TCT (Lord, 1980), al igual que el avión no ha sustituido al autobús como medio de transporte, pero permite llegar a destino en unos márgenes de tiempo impensables para el transporte por carretera.

Así mismo el modelo que constituye la Teoría de Respuesta al Ítem añade dos supuestos a la estimación de las Curvas Características de los Ítems del test.

Unidimensionalidad

La CCI pone en relación las puntuaciones de la variable θ con la probabilidad de responder correctamente un ítem. Para que las inferencias del modelo sean correctas esa probabilidad de acierto debe depender únicamente del valor de θ. Por lo tanto, el modelo asume que los ítems que miden la variable θ son unidimensionales.

Independencia local

De la comprobación de la unidimensionalidad se deriva la asunción de que la respuesta a un ítem no debe influir en la respuesta a los otros. Si existiera esa influencia no se podría sostener que la probabilidad de acierto del ítem depende

únicamente del valor de θ, por lo que no se podría sostener el supuesto de unidimensionalidad.

Modelo

Para escalas del tipo que nos ocupan el modelo más utilizado es el modelo de respuesta graduada de Samejima (1969), mejorado posteriormente por Muraki (1990). Este modelo tiene en cuenta los diferentes puntos de corte entre categorías para cada ítem. Cada uno de estos parámetros de corte expresan el valor de θ a partir del cual es más probable obtener una puntuación de la escala frente a la puntuación ordinal anterior. Su formulación es como se describe a continuación.

$$P_{jk}(\theta)=\frac{1}{1+e^{Da_j(\theta-b_{jk})}}-\frac{1}{1+e^{Da_j(\theta-b_{jk+1})}}$$

Donde a_j es el parámetro de la pendiente para cada ítem y b_{jk} es el parámetro de cada uno de los puntos de corte entre categorías para cada ítem

Función de información

La estimación de la puntuación θ de cada sujeto, al igual que sucede en la Teoría Clásica de los Test, no está exenta de error de medida. Por lo tanto, la Teoría de Respuesta a los Ítems debe de proveerse de algún procedimiento que actúe como indicador de la fiabilidad de la medida.

Sir Ronald Fisher define información como el recíproco de la precisión con la que se puede estimar un parámetro (Muñiz, et al., 2005). La función de información de los ítems y del test aporta datos sobre la precisión con la que se puede estimar la puntuación θ usando el test.

La función de información del test es la expresión gráfica de la curva que pone en relación los diferentes niveles de la variable θ con la información que el test en cada uno de ellos. Lógicamente, la función de información del test es la suma de las funciones de información de sus ítems:

Por si misma, la función de información aporta datos relevantes sobre qué ítems son los mejores, por medir con mayor precisión, para qué niveles de θ.

4.5.2 Análisis de los ítems desde la TRI

El enfoque de la TRI determina las siguientes propiedades como objeto principal de estudio para el análisis de ítems:

- Dificultad.
- Poder discriminativo.
- Bondad de ajuste al modelo.
- Función de información de los ítems y del test.

Tal y como se señaló anteriormente, los modelos clásicos de TRI no resultan útiles para explicar datos obtenidos mediante escalas multicategoriales tipo Likert. Desde finales de los años sesenta del siglo pasado este ha sido uno de los tópicos más recurrentes en la literatura psicométrica, buscando modelos que permitieran el cálculo y la estimación de parámetros para este tipo de ítems (Agresti, 1990; Bock, 1972; Masters, 1982; Mellenbergh, 1995; Molenaar, 1982; Muraki, 1990, 1992; Samejima, 1969; Van der Linden y Hambleton, 1997).

El modelo de respuesta graduada de Samejima (1969) analiza ítems politómicos ordinales, esto implica que, por un lado, la variable de respuesta tiene *k* categorías, siendo $k>2$, y, por otro lado que, existe un orden determinado en las categorías que debe ser conservado. Lo más frecuente es que ese orden sea ascendente, desde una menor a una mayor proximidad a la variable medida. De esta forma, un ítem de este tipo puede segmentarse en una serie de variables de respuesta dicotómica. (Mellenbergh, 1996).

Para cada transición entre alternativas de respuesta se calcula un parámetro de localización distinto, denotado como b, y se estima también un parámetro de discriminación del ítem que será común para los *k-1* parámetros de localización estimados.

Sin embargo, los principales avances del modelo de Samejima no llegaron hasta la década de los noventa. En ese momento Muraki (1990) propuso su modelo de respuesta graduada. En el mantiene los presupuestos del modelo de respuesta graduada original, pero en este caso plantea que los parámetros de localización para

cada categoría de respuesta son equidistantes, con lo que simplifica los cálculos del modelo.

En definitiva, con respecto a la perspectiva de la TRI para los modelos de escalas dicotómicas, el modelo de respuesta graduada estima un parámetro de discriminación *a* común para todas las categorías de respuesta de cada ítem, asimilable al parámetro de los modelos logísticos para escalas dicotómicas. Sin embargo, por contra, el parámetro *b,* no expresa la dificultad del ítem, como sucede en las escalas dicotómicas, sino que es un parámetro de localización que expresa el nivel determinado de θ a partir del cual es más probable que un sujeto escoja una categoría de respuesta frente a su categoría ordinal precedente.

4.5.3 Resultados desde la TRI

El análisis de ítems desde la perspectiva de la TRI realizado utilizando el programa MULTILOG 7.0 (Thissen, Chen, y Bock, 2003), se obtuvo la siguiente función de información para el test completo.

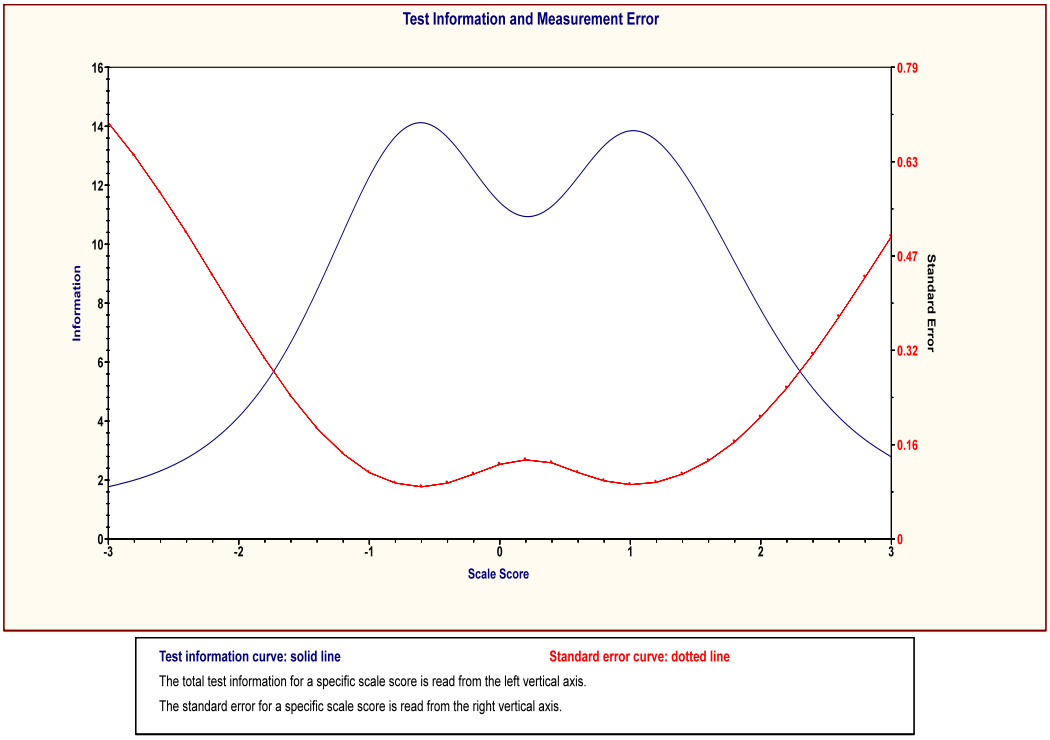


Fig. 23: Función de información de la prueba

En ella puede observarse que la mayor precisión de la medida de la prueba se sitúa en valores medios de la variable θ. El programa calcula además una

Fiabilidad Marginal, entendida, al estilo clásico, como la razón entre la varianza de las puntuaciones verdaderas (que en este caso corresponden con las puntuaciones θ) y la varianza de las puntuaciones empíricas. El resultado tras el análisis es de una fiabilidad marginal para la prueba de evaluación de 0,90.

Para la estimación de los parámetros de la Prueba de Evaluación de conductores desde el enfoque de la TRI se utilizó el programa MULTILOG 7.0 (Thissen, et al., 2003) para poder utilizar un modelo multicategorial. Se aplicó un procedimiento de estimación de máxima verosimilitud marginal (Bock y Lieberman, 1970), el cual viene implementado en el programa. En este procedimiento de estimación se separa la estimación de los parámetros de la estimación de la habilidad mediante una función de verosimilitud que se resuelve integrando sobre la distribución marginal de la variable habilidad. Al no tener carácter cíclico este procedimiento estima todos los parámetros de una vez , por lo que está recomendado para test con pocos ítems, como es este caso.

En la Tabla 23 se resumen los datos obtenidos en la estimación del parámetro *a* y de los diferentes parámetros de localización b_1, b_2 y b_3.

Elemento	*a*	*S.E.*	b_1	b_2	b_3
it 01 EcoRevoluciones	**2,04**	0,09	**-0,70**	**1,14**	**6,82**
it 02 EcoRetarder	**1,85**	0,09	**-0,53**	**1,45**	**6,82**
it 03 EcoFreno	**2,35**	0,09	**-0,78**	**1,44**	**6,34**
it 04 EcoAnticipación	**1,65**	0,07	**-0,78**	**1,88**	**9,17**
it 05 SegControl	**2,09**	0,08	**-0,92**	**1,11**	**7,06**
it 06 SegRetrovisor	**2,00**	0,08	**-0,74**	**1,50**	**7,41**
it 07 SegDistancia	**1,11**	0,06	**-1,94**	**2,26**	**8,54**
it 08 SegVolante	**1,53**	0,06	**-0,97**	**1,20**	**9,03**
it 09 SegSeñalización	**0,95**	0,08	**-1,30**	**1,34**	**11,74**
it 10 TecFluidez	**2,72**	0,10	**-0,83**	**0,96**	**6,01**
it 11 TecMandos	**1,84**	0,08	**-0,17**	**1,45**	**13,90**
it 12 TecCambio	**2,35**	0,09	**-0,46**	**1,00**	**6,63**
it 13 TecManiobra	**3,05**	0,11	**-0,48**	**0,84**	**6,82**

Tabla 23: Parámetros TRI de los ítems

Los valores para el parámetro *a* son bastante buenos, tomando como referencia la escala propuesta por Baker (1985). Casi todos los ítems tienen una alta o muy alta capacidad de discriminación, esto es, por encima de 1,34, siendo el ítem

número 13 (TecManiobra) el de mayor poder discriminativo (**a=3,05**). Únicamente los elementos 7 (SegDistancia) y 9 (SegSeñalización) muestran una capacidad de discriminación *moderada*, según la escala de Baker, por debajo del resto. Los valores son de 1,11 para el ítem 7 y 0,95 para el ítem 9.

Para el parámetro *a,* el error estándar se mantiene bajo, entre 0,06 y 0,11. También lo hace, aunque no se muestra en la tabla, para los parámetros *b*, oscilando entre 0,06 y 0,16.

Los parámetros de localización son bastante homogéneos en la mayoría de los elementos, situándose en el centro de la escala de habilidad. El ítem número 11 (TecCambio) es el que requiere un nivel de habilidad más alto para obtener en el puntuaciones elevadas (Fig. 24)

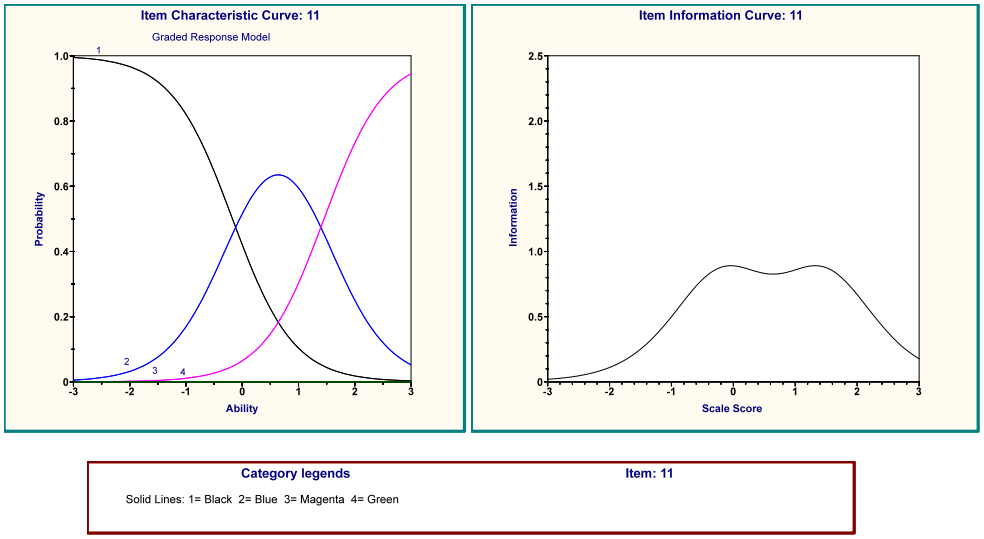


Fig. 24: CCI y Función de información del ítem 11

De nuevo, los elementos 7 y 9 se diferencian del resto. Sus parámetros de localización de la primera categoría b_1 son los más bajos de todos los ítems que conforman la prueba, -1,94 para el ítem 7 y -1,30 para el ítem 9. Así mismo, su función de información es prácticamente plana (Figs. 25 y 26).

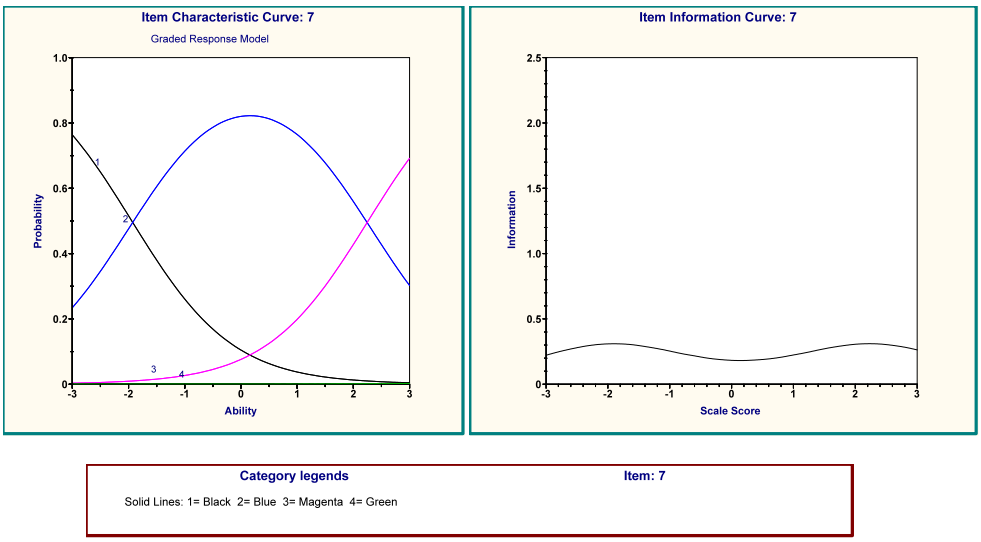


Fig. 25: CCI y Función de información del ítem 7

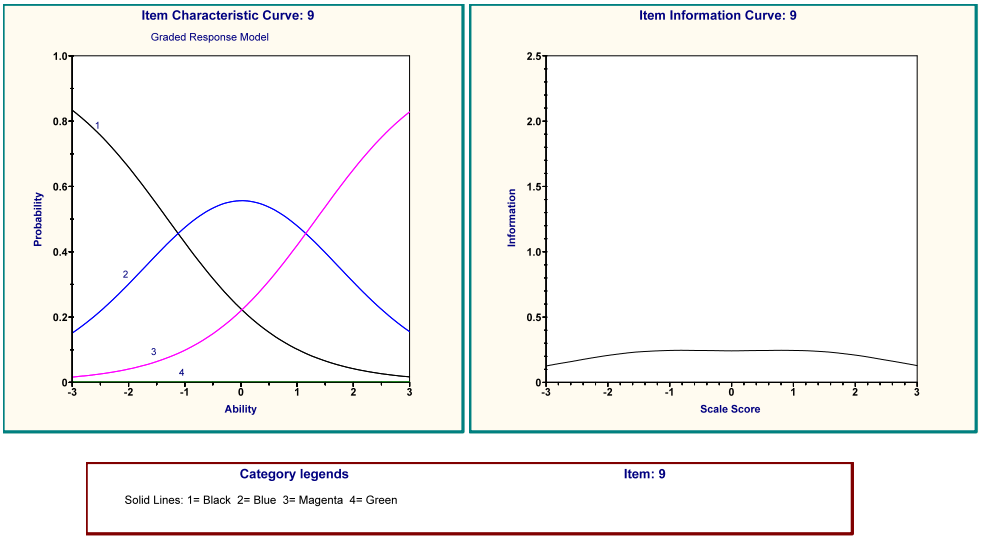


Fig. 26: CCI y Función de información del ítem 9

Por lo que se refiere a la información aportada el elemento número 13 (TecManiobra) es el que mayor información aporta (Fig. 27).

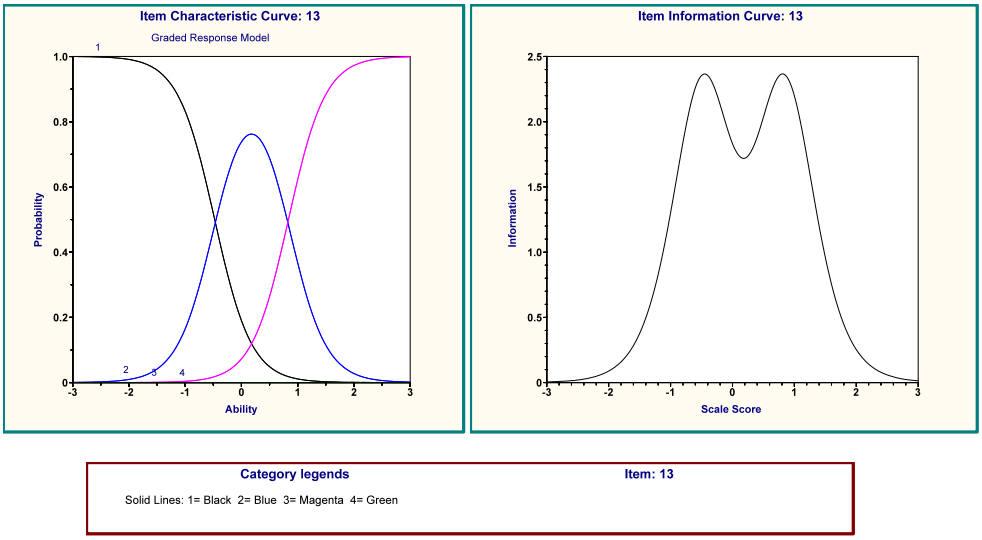


Fig. 27: CCI y Función de información del ítem 13

En el anexo final pueden consultarse todos los datos correspondientes al análisis con MULTILOG y las curvas características de todos los elementos de la prueba.

4.6 Cálculo de la tabla de baremos

En todo proceso de construcción de un test el último paso debe ser, necesariamente, el establecimiento de un baremo general de la prueba y cuantos baremos específicos se consideren necesarios (García Cueto, 1993; Muñiz, 1992, 1998).

La lógica que subyace a este proceso se explica por la necesidad de aportar una información clara sobre la interpretación de las puntuaciones que se obtienen directamente del test. Por ejemplo, decir que un candidato en concreto obtiene una puntuación directa de 21 puntos en la escala de Ansiedad del test NEO PI-R (Costa y McRae, 2008) no constituye, por si misma, una información relevante ni clarificadora. Así aportado el dato no es posible saber si esa puntuación implica que el candidato que ha respondido al test muestra un nivel de ansiedad alto o bajo. La puntuación directa obtenida tras la corrección de un test no es informativa ni fácilmente interpretable por si misma.

Sin embargo, si se argumenta que esa puntuación de 21 supone que el candidato muestra un nivel de ansiedad más alto que el del 60% de la población a la que pertenece, el dato se convierte en información clara y más fácilmente interpretable, sin necesidad de poseer amplios conocimientos matemáticos y de estadística.

Las puntuaciones directas deben, por tanto, ser puestas en relación con la media y desviación típica del grupo normativo al que pertenecen, para que su significado sea interpretable y suficientemente informativo, de tal manera que el test cumpla con su cometido de reportar la situación de un sujeto en concreto en una variable.

El proceso de baremación de un test debe realizarse a partir de la tipificación de las puntuaciones directas de la muestra. Para ello se obtendrán, en primer lugar, las proporciones acumuladas de casos para cada puntuación, esto es, la razón entre la escala centil de las puntuaciones y el valor 100. A partir de este dato se puede calcular fácilmente la puntuación típica normalizada (Z_n) que corresponde a cada una de las puntuaciones directas, ya que conociendo la proporción que queda por debajo de un determinado punto de la distribución se puede acudir a las tablas para comprobar qué puntuación Z_n tiene asociada. La media de esta distribución es 0 y su desviación típica 1 N(0, 1) e indica cuantas desviaciones típicas está por encima, o por debajo, de la media una puntuación dada.

Las puntuaciones típicas normalizadas presentan dos problemas principales que dificultan su interpretación:

1. Pueden asumir valores no enteros.

2. Pueden asumir valores negativos.

Por este motivo y con la intención de superar estas limitaciones, a partir de las puntuaciones Z_n se calculan puntuaciones derivadas normalizadas. Toda vez que las puntuaciones Z_n tienen, como ya se ha mencionado, media 0 y desviación típica 1 se les puede aplicar una transformación lineal para obtener una escala nueva con la media y desviación que interese sin alterar la escala original.

$$Y=bZ_n+a$$

Donde, Y representa la nueva puntuación derivada normalizada; Z_n es el valor de la puntuación típica normalizada; *b* es el valor de la desviación típica que se desee para la nueva escala y *a* representa el valor de la nueva media de la nueva escala resultante de la transformación.

En Psicología son de uso muy frecuente las escalas derivadas normalizadas como, por ejemplo, la escala de Cociente Intelectual, con media 100 y desviación típica 15 CI(100, 15); la escala T, con media 50 y desviación típica 10 T(50, 10) o la escala de Estaninos, con media 5 y desviación típica 2 E(5, 2). Esta última divide a la escala total de puntuaciones en nueve puntos.

Para el establecimiento de los baremos de la Prueba de Evaluación de Conductores se utilizó una escala derivada de decatipos, con media 5 y desviación típica 2'5 N(5, 2'5).

A continuación en la Tabla 24 se puede ver todo el proceso de transformación de las puntuaciones directas obtenidas en el estudio de la muestra general, hasta el establecimiento final de un baremo con puntuaciones en una escala de 0 a 10. Se eligió este formato por su fácil aplicabilidad y comprensión por su relación con el baremo escolar utilizado en España, lo que hace que muy fácilmente entendible que una puntuación de 5 implica un aprobado y la superación de la prueba mientras que puntuaciones en el baremo por debajo de 5 suponen la no superación de la evaluación.

Puntuación	Baremo
0-10	**0**
11-12	**1**
13-15	**2**
16-17	**3**
18-20	**4**
21-22	**5**
23-25	**6**
26-28	**7**
29-30	**8**
31-32	**9**
33-39	**10**

Tabla 24: Baremo de puntuaciones. Muestra general

Compruébese que aquí para una puntuación directa de 23 se obtiene un 6 en el baremo. Recuérdese que 23 es la puntuación que se estableció como punto de corte para determinar la aptitud de un candidato, teniendo en cuenta que el nivel de exigencia era del 60%, por lo tanto el 5 no se considera apto para esta evaluación. Por debajo de esa puntuación se considera que ningún candidato supera la prueba con el nivel mínimo exigido. Nótese que el baremo da una puntuación de 5 a la puntuación directa 22.

Para puntuaciones directas por debajo de 10 se considera que la ejecución en el test indica la carencia total de Habilidad en la Conducción y se les asigna la puntuación 0.

Si no conozco una cosa, la investigaré.

Louis Pasteur (1822-1895)

5 Estudios Específicos

A continuación, en este capítulo se presentan los resultados de diversos estudios complementarios llevados a cabo utilizando los datos de pruebas realizadas con la prueba de evaluación para conductores y orientados a estudiar la relación de los resultados en la prueba con otras variables.

Se analizan, en primer lugar, tres variables poblacionales consideradas de importancia por su relevancia para el uso de la prueba. La primera de ellas será la Edad, la cual se estudia como variable única. Posteriormente, se muestran los resultados de un estudio del efecto de la Experiencia previa con vehículo industrial sobre los resultados en la prueba, primero de manera aislada y después covariando con la edad. La última variable poblacional evaluada presenta los resultados de un estudio de sesgo realizado para verificar si el test tiene un funcionamiento diferencial en función de la variable Sexo.

En segundo lugar se ponen a prueba dos variables personales que tradicionalmente se han relacionado con los índices de accidentalidad en carretera. Se presentan primero los resultados de un estudio realizado para ver la relación de la ejecución en la prueba en función de la capacidad atencional del candidato, para, posteriormente, mostrar los resultados de un segundo estudio en el que se buscan posibles relaciones entre el resultado en la prueba y cinco características de personalidad.

5.1 Edad y resultados en la prueba

5.1.1 Edad y conducción

A lo largo del tiempo son variados los estudios que relacionan la variable demográfica *Edad* con aspectos relacionados con el tráfico y la seguridad en carretera, en Bayam, Liebowitz y Agresti (2005) puede verse un completo meta-análisis de los artículos publicados a este respecto desde el año 1995, y parece evidente que existe una clara relación entre un aumento de la edad y una disminución de las capacidades necesarias para una conducción segura (De Raedt y Ponjaert-Kristoffersen, 2000; Romero, García Sevilla, Martínez Sánchez, Morales, y Sánchez Meca, 1990).

Sin embargo, la gran mayoría de estas investigaciones se centran en lo que se ha dado en denominar *senior driving*, esto es, grupos de conductores cuya edad está por encima de los 65 años. A modo de ejemplo, de los 25 artículos expuestos en el meta-análisis mencionado solo 6 de ellos contemplan conductores por debajo de los 65 años y únicamente 2 usan una muestra de conductores de menos de 60 años. A consecuencia de esto la mayoría de estos estudios se orientan a la investigación de aspectos propios de ese grupo de edad, relacionados con la pérdida de habilidades, la conciencia y percepción del riesgo, efectos diferenciales del accidente por su mayor fragilidad o peores condiciones médicas y el abandono de la conducción.

En general, las mayores diferencias en cuanto a la probabilidad de sufrir un accidente se focalizan en conductores jóvenes, menores de 25 años, y en conductores mayores, por encima de los 65 años, frente a los conductores de

mediana edad (Chipman, MacGregor, Smiley, y Lee-Gosselin, 1993; Clarke, Ward, Bartle, y Truman, 2006, 2010; Li, Braver, y Chen, 2003; Massie, Campbell, y Williams, 1994). McGwin y Brown (1999) determinan que el principal problema para los conductores de más edad está relacionado con la pérdida de capacidades perceptivas y con las dificultades para realizar un juicio correcto, sobre todo en situaciones de presión como, por ejemplo, el cruce de una intersección sin señalizar. Mientras tanto, en los jóvenes la limitación más importante se sitúa en la falta de experiencia, en una menor habilidad al volante y en una mayor dificultad para evaluar las situaciones potencialmente peligrosas. En los conductores de mediana edad parece darse una combinación ideal de experiencia como conductor y capacidad de reacción y procesamiento de la información.

De hecho, en los conductores de edad avanzada se aprecia mayor dificultad para evitar obstáculos, ya que aumenta su distancia mínima de frenado con seguridad (Martin, et al., 2010) y las situaciones que tienen una cierta complejidad, como por ejemplo realizar un adelantamiento o enfrentarse a una intersección sin señalizar, aumentan la carga mental en conductores de mayor edad (Cantin, Lavallière, Simoneau, y Teasdale, 2009).

Sin embargo, los conductores mayores y con más experiencia presentan algunas fortalezas muy positivas para la tarea de la conducción, por ejemplo, evalúan mejor las situaciones potencialmente peligrosas relacionándolas mejor con el entorno del tráfico, existiendo una influencia clave de la experiencia en la evaluación del riesgo potencial (Borowsky, Oron-Gilad, y Parmet, 2009; A. Borowsky, D. Shinar, y T. Oron-Gilad, 2010). Además, se aprecian efectos beneficiosos a consecuencia de la compensación que se da con edades por encima de los 60, resultado de una mayor conciencia de las limitaciones propias del conductor. Por ejemplo, la distancia de seguimiento es mayor y se anticipan con más frecuencia las situaciones del tráfico (Andrews y Westerman, 2012). Por otro lado, ese efecto de autorregulación se traduce en la decisión de dejar de conducir, bien sea definitivamente o bien bajo determinadas condiciones climáticas o a determinadas horas del día (De Raedt y Ponjaert-Kristoffersen, 2000; Donorfio, D'Ambrosio, Coughlin, y Mohyde, 2009)

De todas formas, estos estudios a nivel general no son del todo extrapolables a los conductores profesionales De entrada su exposición al tráfico y al riesgo de accidente es mucho mayor, lo que supone que, al igual que el efecto de la

experiencia es mucho mayor, también lo es el deterioro de las capacidades (Di Milia, et al., 2011) sostienen que este deterioro de capacidades se da en conductores profesionales a partir de los 45 años. Señalan también que el efecto negativo del trabajo a turnos afecta a la conducción y como, además, el ajuste de los ritmos circadianos empeora con la edad.

Ante la escasez actual de conductores profesionales parece sensato pensar que algunos de los que actualmente están en activo y ya tienen una edad avanzada puedan permanecer activos en la profesión por más tiempo, pero ante el hecho de que los conductores profesionales de menos de 27 años y los mayores de 63 sufren un incremento del nivel de accidentalidad (Duke, Guest, y Boggess, 2010) y debido a la limitada cantidad de información disponible para ese segmento de población, parece una decisión difícil de tomar, máxime cuando existe otra dificultad añadida a la hora de determinar la relación entre Edad y Habilidad en la Conducción (Hoggarth, Innes, Dalrymple-Alford, Severinsen, y Jones, 2010).

Teniendo en cuenta que el perfil del conductor profesional de autobuses en España no supera los 65 años y, por normativa, no puede ser inferior a 24 años, parece que la gran mayoría de los conductores se situarían en el perfil del conductor de mediana edad. Para este estudio en concreto la edad media de los conductores es cercana a los 38 años (**M=37,89**) y **DT=8,53**, con una experiencia media de **3,5** años, por lo que no es esperable verificar efectos de la edad en los resultados de la evaluación de conducción.

5.1.2 Estudio y resultados sobre Edad

Para el estudio de los posibles efectos de la edad sobre los resultados de la prueba de evaluación se tomó la muestra general (**N=4425**) y se extrajeron de ella aquellas pruebas de las que se tenía constancia fehaciente del dato correspondiente a la fecha de nacimiento del candidato que realizó la prueba. A partir de ahí la edad fue calculada tomando la diferencia entre la fecha de realización de la prueba y la fecha de nacimiento. De esta forma, el dato de la variable Edad con el que se trabajó a lo largo del estudio corresponde a la edad del conductor exactamente en el momento de realizar la prueba.

Así las cosas, la muestra quedó definitivamente conformada por 3615 pruebas con una media de edad de 38 años (**M=37,39**) y con una desviación típica de 8,53 (**DT=8,53**), tal y como se mostró anteriormente en la Tabla 6 (p. 93).

Para un mejor manejo de la información y para llevar a cabo la técnica de análisis elegida se agrupó a los sujetos por tramos de edad, obteniéndose 7 grupos diferenciados en tramos de 5 años, desde los menores de 25 años hasta los mayores de 50.

Para cada grupo se calculó la puntuación media en la prueba y su desviación típica, tal y como se muestra en la siguiente tabla (Tabla 25).

Estadísticos descriptivos

Variable dependiente:Puntuación Total Test

Tramos de Edad	Media	Desviación típica	N
Menores de 25	**20,00**	**6,261**	**240**
de 25 a 30	**20,99**	**6,305**	**555**
de 30 a 35	**22,37**	**6,151**	**744**
de 35 a 40	**22,28**	**6,200**	**707**
de 40 a 45	**22,23**	**6,173**	**638**
de 45 a 50	**22,31**	**6,233**	**440**
Mayores de 50	**22,85**	**6,259**	**291**
Total	**21,99**	**6,253**	**3615**

Tabla 25: Grupos de Edad

Análisis de Varianza

A partir de esto datos se realizó un Análisis de Varianza (ANOVA) para verificar la posible existencia de diferencias entre los diferentes grupos de edad por lo que se refiere a las puntuaciones en el test.

El ANOVA es una técnica de análisis estadístico que, de la misma forma que otros métodos como la prueba *t*, permite establecer inferencias acerca de las medias. La principal diferencia con otras pruebas de comparación de medias radica en que el ANOVA permite hacer las comparaciones en varias muestras, con diferentes grupos e, incluso, con varias variables independientes, denominadas criterios de clasificación (Tejedor, 1999).

Se trata, por tanto, de una prueba de diferencia de medias que afronta mejor el riesgo de pérdida de información que otras pruebas que tienen un menor número de comparaciones posibles. En otras palabras, se trata de un test de independencia (Agresti y Finlay, 2009) que compara la variable categórica que

caracteriza a los grupos (Edad) con la variable de respuestas cuantitativas que se registra como variable dependiente (Puntuación en la prueba de evaluación).

El método de ANOVA utiliza la extracción de dos estimaciones independientes de la varianza poblacional común. Por un lado, un estimador sesgado de la varianza poblacional que se corresponde con la variabilidad de la media de cada grupo con respecto de la media general, denominada *Varianza Intergrupo*, este estimador solo es insesgado cuando H_0 es verdadera, y por otro, un estimador insesgado que se corresponde con la variabilidad dentro de cada grupo respecto a su media, denominada *Varianza Intragrupo*, este estimador es insesgado independientemente de que H_0 sea verdadera o falsa.

A continuación se muestra la tabla de resultados del ANOVA univariante realizado utilizando el programa SPSS v.17 (Tabla 26).

Variable dependiente:Puntuación Total Test

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
Modelo corregido	**1964,084**[a]	**6**	**327,347**	**8,476**	**,000**	**,014**
Intersección	**1463301,863**	**1**	**1463301,863**	**37891,223**	**,000**	**,913**
EdadCod	**1964,084**	**6**	**327,347**	**8,476**	**,000**	**,014**
Error	**139335,517**	**3608**	**38,618**			
Total	**1889288,000**	**3615**				
Total corregida	**141299,601**	**3614**				

a. R cuadrado = ,014 (R cuadrado corregida = ,012)

Tabla 26: Análisis de varianza para Puntuación en el test por Edad

La primera fila (Modelo corregido) se refiere a todos los efectos del modelo en conjunto y la segunda (Intersección) es la constante del modelo que se utiliza para estimar las medias. A continuación se presentan los efectos principales, en este caso de una única variable independiente que se corresponde con la fila EdadCod. Aquí se informa del efecto individual del factor Edad. Al no haber más factores en el modelo puede comprobarse como en este caso la información de esta fila coincide con la información del Modelo corregido. En la fila Error se contiene la información que corresponde a otras fuentes de variación que no se asocian a los tratamientos. Por último, R cuadrado se obtiene dividiendo la suma de cuadrados del Modelo corregido (1964,084) entre la suma de cuadrados total corregida (141299,601) y, multiplicada por 100, e informa del porcentaje de varianza de la variable dependiente que se explica por los efectos del modelo.

En la tabla se puede comprobar como existen diferencias significativas en las medias de los diferentes grupos por Edad y que la edad explica un 1,4% de la varianza total de la variable Puntuación en el test.

Lo que el ANOVA inicial no define es entre qué grupos se verifican las diferencias, si se dan entre todos o únicamente entre unos grupos en concreto. Es por esto que, en el caso de detectarse diferencias en las medias, se realizan pruebas de comparaciones Post Hoc en las que se puede ver, por pares de grupos, la significación de las diferencias encontradas.

Para este estudio en concreto se utilizó la prueba Scheffe para comparaciones múltiples basadas en las medias observadas y cuyos resultados se muestran a continuación (Tabla 27)

En la tabla puede apreciarse como las diferencias significativas están localizadas en los tramos de edad más jóvenes, por debajo de los 30, mientras que los demás tramos de edad no muestran diferencias significativas entre ellos, por lo que se puede deducir un posible efecto de la experiencia previa con vehículo pesado que se estudiará a continuación.

Puntuación Total Test
Scheffe

(I)Tramos de Edad	(J)Tramos de Edad	Diferencia de medias (I-J)	Error típ.	Sig.	Intervalo de confianza 95% Límite inferior	Límite superior
Menores de 25	de 25 a 30	-,99	,480	,639	-2,70	,71
	de 30 a 35	-2,37*	,461	,000	-4,00	-,73
	de 35 a 40	-2,28*	,464	,000	-3,93	-,63
	de 40 a 45	-2,23*	,471	,001	-3,90	-,55
	de 45 a 50	-2,31*	,499	,002	-4,08	-,54
	Mayores de 50	-2,85*	,542	,000	-4,77	-,92
de 25 a 30	Menores de 25	,99	,480	,639	-,71	2,70
	de 30 a 35	-1,37*	,349	,017	-2,61	-,14
	de 35 a 40	-1,29*	,352	,037	-2,54	-,04
	de 40 a 45	-1,23	,361	,070	-2,51	,05
	de 45 a 50	-1,32	,397	,086	-2,73	,09
	Mayores de 50	-1,86*	,450	,009	-3,45	-,26
de 30 a 35	Menores de 25	2,37*	,461	,000	,73	4,00
	de 25 a 30	1,37*	,349	,017	,14	2,61
	de 35 a 40	,08	,326	1,000	-1,08	1,24
	de 40 a 45	,14	,335	1,000	-1,05	1,33
	de 45 a 50	,05	,374	1,000	-1,28	1,38
	Mayores de 50	-,48	,430	,974	-2,01	1,04
de 35 a 40	Menores de 25	2,28*	,464	,000	,63	3,93
	de 25 a 30	1,29*	,352	,037	,04	2,54
	de 30 a 35	-,08	,326	1,000	-1,24	1,08
	de 40 a 45	,06	,339	1,000	-1,15	1,26
	de 45 a 50	-,03	,377	1,000	-1,37	1,31
	Mayores de 50	-,57	,433	,944	-2,10	,97
de 40 a 45	Menores de 25	2,23*	,471	,001	,55	3,90
	de 25 a 30	1,23	,361	,070	-,05	2,51
	de 30 a 35	-,14	,335	1,000	-1,33	1,05
	de 35 a 40	-,06	,339	1,000	-1,26	1,15
	de 45 a 50	-,09	,385	1,000	-1,46	1,28
	Mayores de 50	-,62	,440	,919	-2,18	,94
de 45 a 50	Menores de 25	2,31*	,499	,002	,54	4,08
	de 25 a 30	1,32	,397	,086	-,09	2,73
	de 30 a 35	-,05	,374	1,000	-1,38	1,28
	de 35 a 40	,03	,377	1,000	-1,31	1,37
	de 40 a 45	,09	,385	1,000	-1,28	1,46
	Mayores de 50	-,54	,470	,972	-2,20	1,13
Mayores de 50	Menores de 25	2,85*	,542	,000	,92	4,77
	de 25 a 30	1,86*	,450	,009	,26	3,45
	de 30 a 35	,48	,430	,974	-1,04	2,01
	de 35 a 40	,57	,433	,944	-,97	2,10
	de 40 a 45	,62	,440	,919	-,94	2,18
	de 45 a 50	,54	,470	,972	-1,13	2,20

Basadas en las medias observadas.
El término de error es la media cuadrática(Error) = 38,618.

*. La diferencia de medias es significativa al nivel ,05.

Tabla 27: Comparaciones Post Hoc. Scheffe para Edad

Por último, en la figura 28 puede verse la representación gráfica de las medias marginales estimadas, donde se observa como a medida que aumentan los tramos de edad sube la media en la puntuación de la prueba.

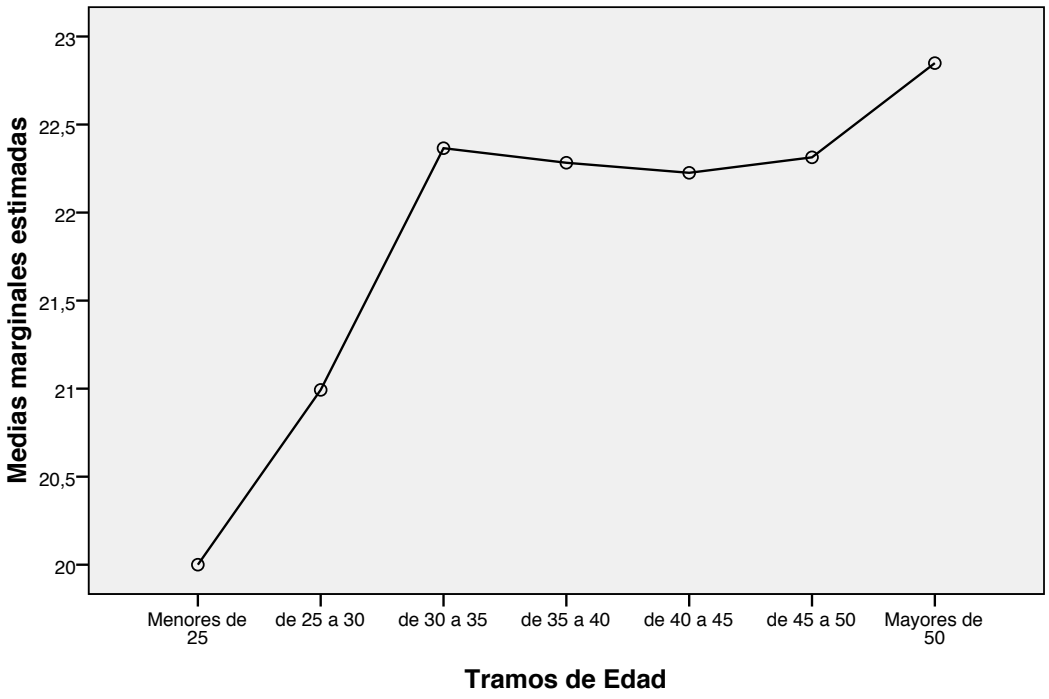


Fig. 28: Gráfico de medias marginales estimadas. ANOVA Edad

Cálculo del tamaño del efecto

Cuando se lleva a cabo un estudio de contraste estadístico de medias, como el que se ha realizado aquí, mediante la realización de un Análisis de Varianza, la American Psychological Association (APA, 2001) recomienda que se aporte información complementaria al contraste mostrando al lector el valor del tamaño del efecto (*effect size*).

Esto se debe a que el hecho de que exista una diferencia estadísticamente significativa no implica necesariamente que exista una diferencia grande, ni tampoco necesariamente importante. El valor del estadístico de contraste no es dependencia única del tamaño de la diferencia, sino que también está afectado por la variabilidad y, por encima de todo, del tamaño de la muestra.

La consecuencia de esto es que, ateniéndose únicamente a la significación estadística, con una muestra lo suficientemente grande puede probarse casi cualquier cosa, cuando, en realidad, una diferencia estadísticamente significativa puede no ser relevante a efectos práctico y viceversa. El profesor Morales Vallejo, en

un magnífico artículo, profundiza en las limitaciones de la inferencia estadística por contraste de medias y, así mismo, aporta una extensa revisión bibliográfica sobre este particular para el lector interesado (Morales Vallejo, 2010).

El tamaño del efecto permite conocer la magnitud de la diferencia y, por consiguiente, la relevancia que tiene la diferencia encontrada (Morales Vallejo, 2010).

Para el cálculo del tamaño del efecto en el presente estudio se utilizó el procedimiento de Hedges y Olkin utilizando la fórmula que se presenta a continuación:

$$g=\frac{\dot{X}_1-\dot{X}_2}{\sigma} \quad \text{donde} \quad \sigma=\sqrt{\frac{(N_1-1)\sigma^2_{n-1(1)}+(N_2-1)\sigma^2_{n-1(2)}}{N_1+N_2-2}}$$

En el caso de tener *K* grupos, como sucede en el Análisis de Varianza, el denominador puede sustituirse por la raíz cuadrada de los cuadrados medios Intragrupo y quedaría formulado como se describe a continuación:

$$g=\frac{|M_i-M_k|}{\sigma}$$

Aquí la diferencia de medias entre los pares de grupos se presenta en valor absoluto aunque, en realidad, el signo es independiente de la magnitud de la diferencia y solo viene determinado por el orden en el que se dispongan las medias de los grupos.

Una vez calculado el valor del tamaño del efecto para una diferencia de medias concreta resta realizar la interpretación de ese valor. No existe un consenso al respecto, aunque si se han aportado orientaciones que son aceptadas y de uso común. A este respecto quien más ha aportado es Cohen (1988) quien proporciona una orientación clara para interpretar el valor de d y que se puede resumir de la siguiente manera:

- *d=0,20* tamaño pequeño.
- *d=0,50* tamaño moderado.
- *d=0,80* tamaño grande.

Esto es, con valores alrededor de 0,20 el tamaño del efecto debe interpretarse como pequeño, con valores rondando 0,50 como moderado y en valores cercanos a 0,80 debe pensarse que el tamaño del efecto puede considerarse como grande.

Un estudio de Bloom y Lipsey (Bloom y Lipsey, 2004), en el que se describe un meta-análisis realizado por el segundo autor, señala que, tras revisar 6.700 estudios, la distribución de los tamaños del efecto tiene su tercio inferior entre 0,00 y 0,32, el tercio central ente 0,33 y 0,55 y el tercio superior entre 0,56 y 1,26.

A partir de ambas informaciones para esta tesis doctoral se ha interpretado el tamaño del efecto de acuerdo a la siguiente lista:

- $d \leq 0,30$ tamaño pequeño.
- $0,30 < d < 0,80$ tamaño moderado.
- $d \geq 0,80$ tamaño grande.

A continuación se muestra la matriz de datos que recoge los resultados del cálculo del tamaño del efecto para las comparaciones por pares de los grupos de edad del estudio[9] (Tabla 28).

[9] El cálculo no está programado en SPSS. Para realizarlo se utilizó una hoja Excel programada a tal efecto por Ignacio Pedrosa, de la Universidad de Oviedo.

Grupos	<25	25-30	30-35	35-40	40-45	45-50	>50
<25	0						
25-30	0,16	0					
30-35	**0,38**	0,22	0				
35-40	**0,37**	0,21	0,01	0			
40-45	**0,36**	0,20	0,02	0,01	0		
45-50	**0,37**	0,21	0,01	0,00	0,01	0	
>50	**0,46**	0,30	0,08	0,09	0,10	0,09	0

Grupos	<25	25-30	30-35	35-40	40-45	45-50	>50
<25	0						
25-30	Pequeño	0					
30-35	**Moderado**	Pequeño	0				
35-40	**Moderado**	Pequeño	Pequeño	0			
40-45	**Moderado**	Pequeño	Pequeño	Pequeño	0		
45-50	**Moderado**	Pequeño	Pequeño	Pequeño	Pequeño	0	
>50	**Moderado**	Pequeño	Pequeño	Pequeño	Pequeño	Pequeño	0

Tabla 28: Matriz de Tamaño del efecto (*g* de Hedges para EDAD-PUNTUACIÓN)

Puede comprobarse que el tamaño del efecto es pequeño para todos los grupos salvo para las diferencias del grupo de menos de 25 años que muestra unas diferencias de efecto moderado con respecto a los grupos por encima de 30 años.

De nuevo las principales diferencias se localizan en el grupo de menor edad y menos experiencia, lo que refuerza la idea de que puede existir un efecto de la experiencia previa con vehículo en la puntuación obtenida en el test.

Por último, se señalará que el valor aportado en la tabla de ANOVA (Tabla 28, p. 177) y que corresponde a la columna *Eta cuadrado parcial*, con valor **0,014** para la fila asociada a los tratamientos (EdadCod), se interpreta como proporción de varianza explicada y sirve como indicador de la potencia asociada al contraste realizado para ese efecto. Se calcula dividiendo el producto del estadístico *F* y los grados de libertad del efecto entre la suma de ese mismo producto con los grados de libertad del error.

5.2 Efecto de la experiencia sobre los resultados en la prueba

5.2.1 Experiencia y conducción

Tanto la Edad como la Experiencia en conducción son dos predictores clásicos utilizados con mucha frecuencia para el estudio de la accidentalidad y el riesgo de tener un accidente. Sin embargo, ambos suelen presentarse de una forma muy intercorrelacionada (Maycock, Lester, y Lockwood, 1996), lo que los convierte en dos aspectos muy difíciles de separar cuando se realiza una investigación (Brown, 1982; Mayhew y Simpson, 1990).

Aún así, existe abundante literatura en la que se ha estudiado el efecto de la experiencia sobre los conductores. Por ejemplo, Kass, Cole y Stanny (2007) han estudiado los efectos de la experiencia sobre la conducción y concluyen que aquellos conductores con menor experiencia previa cometen más infracciones y, sobre todo, muestran un menor nivel de lo que los autores denominan *alerta situacional*. Esto implica que los conductores menos experimentados presentan mayores dificultades para identificar las pistas y los estímulos relevantes del entorno, además realizan peor la integración de la información obtenida. Como consecuencia el modelo de representación situacional que construye el conductor novel es deficiente, lo que le impide anticipar los sucesos futuros de manera eficiente, todo lo contrario que le pasa a un conductor con más experiencia, el cual no muestra un nivel tan bajo de *alerta situacional*.

Siguiendo esa misma línea de investigación Underwood et alt. (2003) detectó la existencia de una relación entre la experiencia al volante y la amplitud del campo perceptivo efectivo. Aquellos conductores con más experiencia gestionan mejor la información presente en su foco atencional y, no solo eso, además son capaces de recordar más eventos potencialmente peligrosos, frente a aquellos conductores con menos experiencia. Parece como si, de alguna forma, la experiencia previa al volante optimizase el procesamiento de la información visual e hiciera más eficiente la gestión de los diferentes focos atencionales.

A este respecto Konstantopoulos et alt. (2010), ante el hecho de que la falta de experiencia parece ser una de los principales factores implicados en la mayor accidentalidad entre los jóvenes (Chapman y Underwood, 1998; Gregersen y Bjurulf, 1996; Neyens y Boyle, 2008), trataron de relacionar, en sus investigaciones, la inexperiencia con patrones de atención visual deficientes. Encontraron que conductores con experiencia muestran una mayor ratio de objetos detectados, requieren menor tiempo de procesamiento de la información y poseen un escaneo situacional más amplio. Estos investigadores prestaron especial atención a la influencia que sobre esta capacidad de escaneo tiene la rutina de revisión de los espejos retrovisores. Determinaron que, en los conductores experimentados, la frecuencia de observación de los espejos es mayor, proporcionándoles de esta manera una información más completa del entorno de conducción.

Otra línea de estudio clásico es la revisión de la relación entre la experiencia al volante y la percepción del riesgo. En general las investigaciones en este área se dirigen a tratar de explicar la disminución del riesgo potencial de accidente transcurridos unos meses de experiencia al volante. Sagberg y Bjørnskau (2006) propusieron que este descenso podría estar relacionado con una mejora en la capacidad del conductor para percibir, identificar y procesar las situaciones de riesgo potencial. Lamentablemente, su estudio no obtiene resultados significativos que sustenten una relación entre la disminución del riesgo de accidente y una mejora en la capacidad de percepción del riesgo vía una mayor experiencia al volante.

Sin embargo, parece evidente que la práctica repetida y la experiencia de diferentes situaciones al volante dotan al conductor de una serie de habilidades prácticas que mejoran su ejecución general. Borowsky et alt. (2010) determinaron que, precisamente, es la experiencia la que interviene en la mejora de la percepción

del riesgo al mejorar la capacidad del conductor de evaluar las diferentes situaciones como potencialmente peligrosas. Esto contribuye a que la dirección de la mirada se dirija, con mayor efectividad, a aquellos lugares que se han identificado previamente y por efecto de la experiencia como potencialmente peligrosos.

Últimamente, Scialfa et alt. (2012) han propuesto un test para evaluar esta capacidad perceptiva del riesgo y utilizándolo en una primera investigación han encontrado que la muestra de conductores jóvenes sin experiencia que utilizaron evalúa los riesgos como potencialmente menos peligrosos, diferenciándose no solo de la muestra de conductores de más edad y más experimentados, sino también de la muestra de conductores jóvenes que muestran cierta experiencia previa en la conducción.

Por último, a pesar de no ser un tema demasiado estudiado en la literatura científica relacionada, se comentarán aquí, por su especial relevancia, un estudio de Harrison y Fillmore (2005) en el que se pone en relación el consumo de alcohol con la experiencia previa al volante del conductor que ha bebido. En el estudio mencionado se pone de manifiesto que, como por otra parte era de esperar, niveles altos de alcohol en sangre reducen significativamente las habilidades y las capacidades para conducir. Sin embargo, se verifica además que esta merma de capacidades es mucho más importante en aquellos conductores que tienen menos experiencia y por tanto con menor sustrato previo en cuanto a habilidades y capacidades para la tarea de la conducción.

Como ha podido comprobarse en este breve recorrido, existe cumplida literatura con respecto al tema de la experiencia en la conducción, pero la mayoría de lo publicado tiene como principal objeto de estudio al conductor de un vehículo privado y poco se ha hecho al respecto para verificar si los resultados obtenidos con estos pueden ser extrapolados y generalizados a conductores profesionales, ya sea de autobús o de camión.

Con este último objetivo en mente Dorn y af Wåhlberg (2008) realizaron un estudio, a partir de los datos de una empresa de autobuses de Inglaterra, en el que encontraron resultados similares, para los conductores de autobús, a los obtenidos para los conductores de vehículos particulares. En general, para la muestra estudiada, el riesgo de accidente en conductores profesionales de autobús desciende con la edad y cuanto mayor es la experiencia previa de estos conductores con vehículos similares. De ninguna manera la antigüedad del permiso de conducir

es indicativo de mayor experiencia, ni se relaciona con un menor riesgo de accidentalidad. Estos autores también encuentran que la experiencia tiene un efecto más fuerte sobre la accidentalidad sobre todo durante el primer año de trabajo, mientras que la edad muestra una asociación en forma de *U*, esto es, los más jóvenes y los mayores son los que muestran mayor riesgo de accidente, mientras que este disminuye para los conductores de mediana edad.

Como complemento al estudio realizado en esta tesis doctoral se estudió la relación de la experiencia previa de un grupo de conductores con los resultados obtenidos en la prueba.

5.2.2 Estudio y resultados sobre Experiencia

Para realizar el estudio del posible efecto de la experiencia sobre las puntuaciones de la prueba se tomó una submuestra de 477 sujetos, proveniente de la muestra total (N=4425).

El procedimiento de extracción de la muestra se realizó mediante un muestreo aleatorio. Inicialmente se eligieron aleatoriamente 12 números del 0 al 99. A partir de ese punto se ordenó el listado de pruebas por fecha de aplicación, desde la más antigua hasta la más reciente, y se extrajeron 12 pruebas por centena cuya posición correspondía con los números obtenidos, previamente, al azar. De las 531 pruebas extraídas definitivamente se eliminaron aquellas que, o bien tenían valores missing, o bien de las que no se disponía de información sobre la experiencia previa del candidato que había realizado la prueba. Al finalizar este proceso quedaron las 477 pruebas con las que se trabajó finalmente.

Para este estudio en concreto no se diferenció el tipo de experiencia entre camión de más de 3’5tn o autobús. Se tomó toda la experiencia previa del candidato como un todo, teniendo en cuenta los años de trabajo profesional efectivo como conductor de un vehículo pesado independientemente del tipo de vehículo o recorrido que realizara.

Para el estudio estadístico se organizó la muestra seleccionada en 4 grupos diferenciados por el total de años de experiencia previa con vehículo pesado de los candidatos que realizaron la prueba. La Tabla 29 muestra la distribución definitiva de los grupos.

Factores inter-sujetos

		Etiqueta del valor	N
Experiencia con vehículo pesado	1	**Menos de 1 año**	**222**
	2	**1 a 3 años**	**84**
	3	**3 a 5 años**	**45**
	4	**5 o más años**	**126**

Tabla 29: Distribución de los grupos por experiencia profesional con vehículo pesado

Utilizando esta distribución de muestra se realizó un análisis de varianza univariante con el objeto de estudiar si existen diferencias en las puntuaciones obtenidas por los diferentes grupos en la prueba de evaluación que pudieran responder a un efecto de la experiencia previa con vehículo pesado.

A continuación se muestra la tabla resumen de análisis de varianza con los resultados obtenidos para la variable independiente Experiencia en conducción, codificada en la tabla como ExpCod, sobre la variable dependiente Resultado total de la prueba (Tabla 30).

Pruebas de los efectos inter-sujetos

Variable dependiente:Puntuación Total Test

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
Modelo corregido	**1671,391**[a]	**3**	**557,130**	**15,626**	**,000**	**,090**
Intersección	**178283,420**	**1**	**178283,420**	**5000,456**	**,000**	**,914**
ExpCod	**1671,391**	**3**	**557,130**	**15,626**	**,000**	**,090**
Error	**16864,072**	**473**	**35,653**			
Total	**250108,000**	**477**				
Total corregida	**18535,463**	**476**				

a. R cuadrado = ,090 (R cuadrado corregida = ,084)

Tabla 30: Análisis de varianza para Puntuación en el test por Experiencia

Puede comprobarse como en la fila correspondiente a los efectos principales del modelo (ExpCod) el análisis da como resultado una diferencia de medias significativa (*F=15,626*).

Como se realizó anteriormente, para determinar entre cuáles de los grupos se dan las diferencias significativas se llevó a cabo un contraste Post Hoc. En este caso se utilizó como estadístico la prueba C de Dunnett ya que el contraste de Levene, que pone a prueba la igualdad de varianzas, da como resultado no homoscedasticidad[10] y esta prueba de Dunnett se realiza asumiendo varianzas no iguales, por lo que resulta más apropiado su uso en esta situación.

[10] Puede verse la tabla de cálculo del estadístico en el anexo correspondiente a este análisis de varianza.

La Tabla 31 muestra los resultados de las comparaciones Post Hoc y los valores del estadístico C de Dunnett para las diferencias de medias entre los grupos. Marcados con un asterisco aparecen aquellos valores de diferencia de medias que resultan ser estadísticamente significativos.

Comparaciones múltiples

Variable dependiente:Puntuación Total Test

C de Dunnett	Menos de 1 año	1 a 3 años	**-2,01***	**,706**
		3 a 5 años	**-4,21***	**,855**
		5 o más años	**-4,14***	**,646**
	1 a 3 años	Menos de 1 año	**2,01***	**,706**
		3 a 5 años	**-2,20**	**,887**
		5 o más años	**-2,12***	**,688**
	3 a 5 años	Menos de 1 año	**4,21***	**,855**
		1 a 3 años	**2,20**	**,887**
		5 o más años	**,07**	**,841**
	5 o más años	Menos de 1 año	**4,14***	**,646**
		1 a 3 años	**2,12***	**,688**
		3 a 5 años	**-,07**	**,841**

Basadas en las medias observadas.
El término de error es la media cuadrática(Error) = 35,653.

*. La diferencia de medias es significativa al nivel ,05.

Tabla 31: Comparaciones Post Hoc. C de Dunnett para Experiencia

La tabla muestra como el grupo de los candidatos con menor experiencia previa con vehículo (Menos de 1 año) tiene diferencias estadísticamente significativas con todos los demás grupos ($p<0.05$). Para el resto de los grupos las diferencias se dan cuando hay una distancia significativa en los años de experiencia, esto es, el grupo de *1 a 3 años* de experiencia muestra diferencias significativas con el grupo de *5 o más años*, sin embargo no muestra esas diferencias con el grupo de *3 a 5 años*, con el que le separan menos años.

El gráfico de medias marginales obtenido a partir del análisis de varianza efectuado deja ver como a medida que los grupos acreditan más años de experiencia las puntuaciones medias en la prueba de evaluación van en aumento, estabilizándose a partir del grupo de *3 a 5 años*. (Fig. 29)

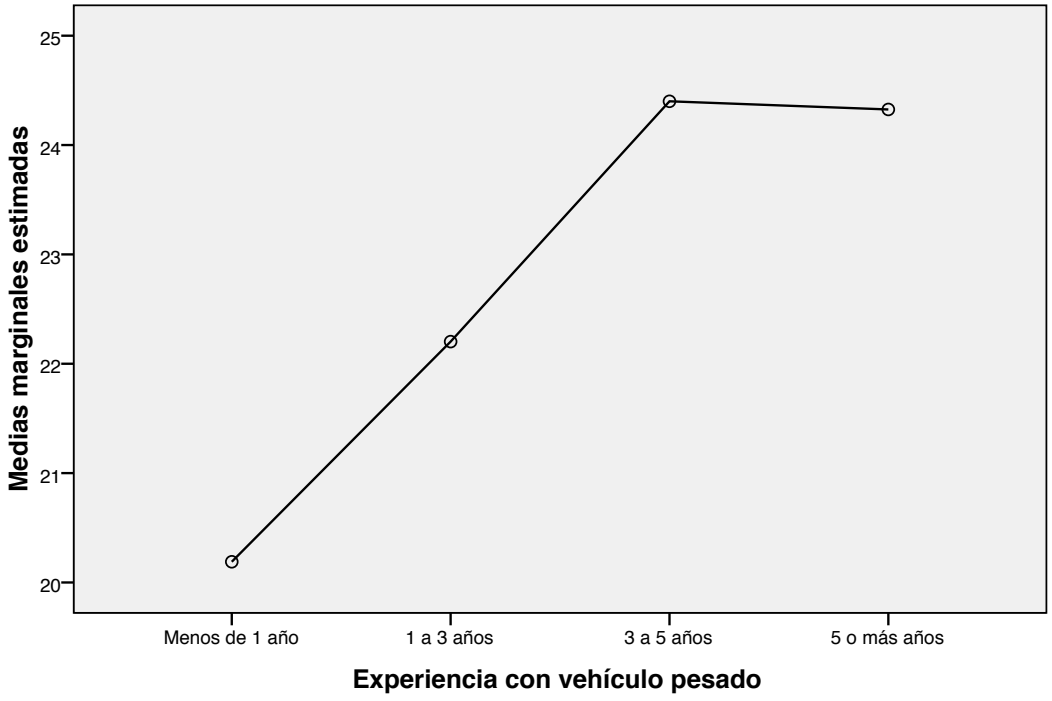


Fig. 29: Gráfico de medias marginales estimadas. ANOVA Experiencia

Por último, se calculó el tamaño del efecto para las diferencias de medias estimadas en el análisis de varianza, utilizando el mismo método que fue utilizado anteriormente para el análisis realizado sobre la variable Edad. (Tabla 32)

Grupos por Experiencia	**menos de 1 año**	**1 a 3 años**	**3 a 5 años**	**5 o más años**
menos de 1 año	0			
1 a 3 años	**0,31**	0		
3 a 5 años	**0,63**	**0,46**	0	
5 o más años	**0,66**	**0,44**	0,02	0

Grupos por Experiencia	**menos de 1 año**	**1 a 3 años**	**3 a 5 años**	**5 o más años**
menos de 1 año	0			
1 a 3 años	**Moderado**	0		
3 a 5 años	**Moderado**	**Moderado**	0	
5 o más años	**Moderado**	**Moderado**	Pequeño	0

Tabla 32: Matriz de Tamaño del efecto (g de Hedges para EXPERIENCIA-PUNTUACIÓN)

En la matriz resultante se observa como todas las diferencias muestran un tamaño del efecto moderado con la única excepción de la diferencia de medias

entre el grupo de *3 a 5 años* y el de *5 o más años*, que resulta ser, además, una diferencia no significativa. Cabe destacar también que la diferencia entre el grupo *3 a 5 años* y el grupo *1 a 3* tiene un tamaño del efecto moderado (*g=0,46*), sin embargo la diferencia de medias estimada en las comparaciones Post Hoc no resulta ser estadísticamente significativa.

Por último, debe señalarse también que la potencia observada para los efectos inter-sujetos no resulta estadísticamente significativa y que el valor de Eta cuadrado parcial (*0,09*) implica que el modelo, teniendo en cuenta la experiencia previa en conducción, explica un 9% de la varianza total de las puntuaciones del test.

Tras comprobar el efecto que la experiencia tiene sobre las puntuaciones finales observadas en la prueba de evaluación se decidió repetir el análisis de varianza que se había efectuado sobre los grupos de la variable Edad, pero, en esta ocasión, controlando el posible efecto de la experiencia sobre la edad.

Para ello se llevó a cabo un análisis de varianza univariante introduciendo la Puntuación total en el test como variable dependiente, la Edad como variable independiente de efectos fijos (introduciendo todos los grupos de edad posibles) y la Experiencia como covariable. Se obtuvo el siguiente cuadro de ANOVA. (Tabla 33)

Pruebas de los efectos inter-sujetos

Variable dependiente:Puntuación Total Test

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
Modelo corregido	**1519,291**[a]	**7**	**217,042**	**6,011**	**,000**	**,083**
Intersección	**87925,210**	**1**	**87925,210**	**2435,170**	**,000**	**,839**
ExpTot	**899,941**	**1**	**899,941**	**24,925**	**,000**	**,051**
EdadCod	**532,686**	**6**	**88,781**	**2,459**	**,024**	**,031**
Error	**16861,686**	**467**	**36,106**			
Total	**249647,000**	**475**				
Total corregida	**18380,977**	**474**				

a. R cuadrado = ,083 (R cuadrado corregida = ,069)

Tabla 33: Análisis de varianza para Puntuación en el test por Edad con Experiencia como covariable

En la tabla se puede verificar que la diferencia de medias es significativa, sin embargo las comparaciones por pares efectuadas posteriormente, mediante el ajuste de Bonferroni, no da como resultado ningún intervalo de confianza significativo para la diferencia de medias por pares. (Tabla 34)

Comparaciones por pares

Variable dependiente:Puntuación Total Test

(I)Tramos de Edad	(J)Tramos de Edad	Diferencia de medias (I-J)	Error típ.	Sig.[a]	Intervalo de confianza al 95 % para la diferencia[a] Límite inferior	Límite superior
Menores de 25	de 25 a 30	-1,716	1,107	1,000	-5,097	1,664
	de 30 a 35	-3,117	1,104	,104	-6,489	,255
	de 35 a 40	-1,516	1,109	1,000	-4,902	1,871
	de 40 a 45	-1,981	1,180	1,000	-5,586	1,623
	de 45 a 50	,037	1,326	1,000	-4,015	4,088
	Mayores de 50	,509	1,885	1,000	-5,251	6,269
de 25 a 30	Menores de 25	1,716	1,107	1,000	-1,664	5,097
	de 30 a 35	-1,401	,869	1,000	-4,056	1,255
	de 35 a 40	,201	,874	1,000	-2,469	2,871
	de 40 a 45	-,265	,958	1,000	-3,191	2,662
	de 45 a 50	1,753	1,134	1,000	-1,712	5,218
	Mayores de 50	2,225	1,751	1,000	-3,123	7,574
de 30 a 35	Menores de 25	3,117	1,104	,104	-,255	6,489
	de 25 a 30	1,401	,869	1,000	-1,255	4,056
	de 35 a 40	1,601	,850	1,000	-,997	4,199
	de 40 a 45	1,136	,920	1,000	-1,675	3,946
	de 45 a 50	3,154	1,106	,095	-,224	6,532
	Mayores de 50	3,626	1,716	,738	-1,617	8,869
de 35 a 40	Menores de 25	1,516	1,109	1,000	-1,871	4,902
	de 25 a 30	-,201	,874	1,000	-2,871	2,469
	de 30 a 35	-1,601	,850	1,000	-4,199	,997
	de 40 a 45	-,465	,914	1,000	-3,257	2,326
	de 45 a 50	1,552	1,102	1,000	-1,813	4,918
	Mayores de 50	2,025	1,708	1,000	-3,193	7,243
de 40 a 45	Menores de 25	1,981	1,180	1,000	-1,623	5,586
	de 25 a 30	,265	,958	1,000	-2,662	3,191
	de 30 a 35	-1,136	,920	1,000	-3,946	1,675
	de 35 a 40	,465	,914	1,000	-2,326	3,257
	de 45 a 50	2,018	1,137	1,000	-1,455	5,490
	Mayores de 50	2,490	1,707	1,000	-2,725	7,705
de 45 a 50	Menores de 25	-,037	1,326	1,000	-4,088	4,015
	de 25 a 30	-1,753	1,134	1,000	-5,218	1,712
	de 30 a 35	-3,154	1,106	,095	-6,532	,224
	de 35 a 40	-1,552	1,102	1,000	-4,918	1,813
	de 40 a 45	-2,018	1,137	1,000	-5,490	1,455
	Mayores de 50	,472	1,822	1,000	-5,095	6,040
Mayores de 50	Menores de 25	-,509	1,885	1,000	-6,269	5,251
	de 25 a 30	-2,225	1,751	1,000	-7,574	3,123
	de 30 a 35	-3,626	1,716	,738	-8,869	1,617
	de 35 a 40	-2,025	1,708	1,000	-7,243	3,193
	de 40 a 45	-2,490	1,707	1,000	-7,705	2,725
	de 45 a 50	-,472	1,822	1,000	-6,040	5,095

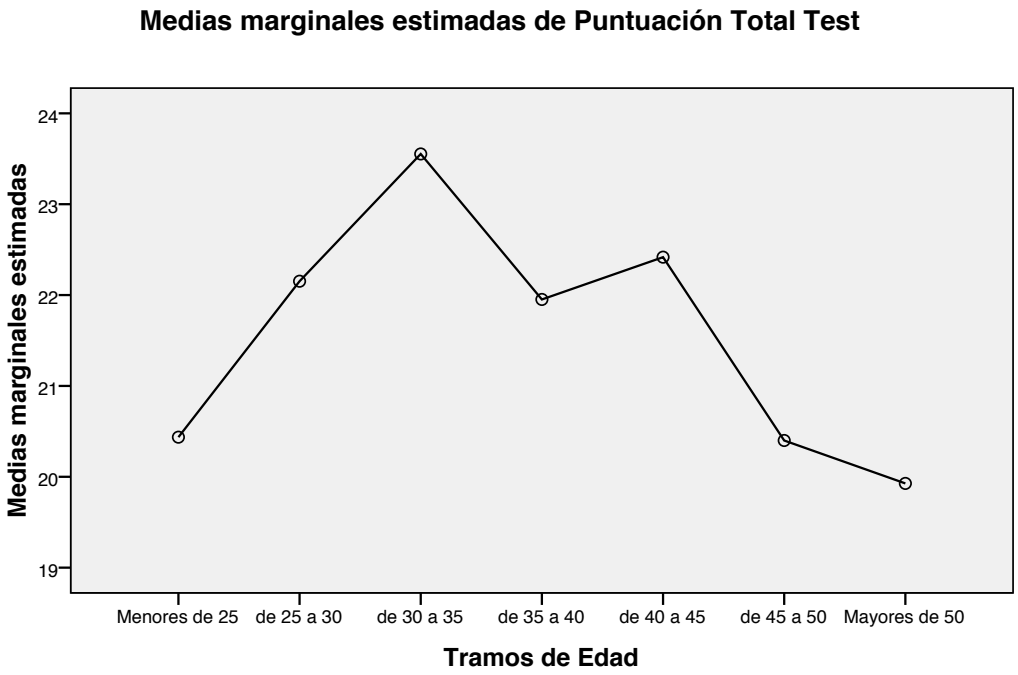
Basadas en las medias marginales estimadas.

a. Ajuste para comparaciones múltiples: Bonferroni.

Tabla 34: Comparaciones por pares. Ajuste Bonferroni para Edad con Experiencia como covariable

En el gráfico de medias marginales estimadas puede verse como, tras controlar el efecto de la covariable Experiencia, las medias para la variable edad cambian con respecto a lo expuesto en la figura 31 (p. 185) y se observa un efecto

de *U* invertida, donde los grupos de candidatos más jóvenes y aquellos de más edad, son los que obtienen peores puntuaciones. (Fig. 30)



Las covariables que aparecen en el modelo se evalúan en los siguiente valores: Años Experiencia = 3,56

Fig. 30: Gráfico de medias marginales estimadas. ANOVA Edad con Experiencia como covariable

5.3 Sexo y resultados en la prueba: Estudio de sesgo

5.3.1 Sexo y conducción

Desde hace ya bastante tiempo el estudio de las diferencias entre hombres y mujeres, en cuanto a su conducción, gestión de la información al volante y su distinta implicación en los accidentes de tráfico, es un lugar común en la literatura científica especializada.

A finales de los años setenta un estudio realizado en Inglaterra no encontró diferencias significativas en cuanto a riesgo de padecer un accidente de tráfico entre hombres y mujeres, pero si marcó una línea de pensamiento y trabajo que actualmente se encuentra muy desarrollada (Storie, 1977). Lo que este informe revela es que, a pesar de no haber encontrado diferencias significativas por sexo en cuanto a riesgo de accidente, los hombres tenían mayor probabilidad de tener un accidente motivado por algún tipo de transgresión o infracción de la norma, mientras que en el caso de las mujeres estas tenían más probabilidad de que el accidente estuviera motivado por errores de tipo perceptivo.

Este planteamiento se ha visto ratificado por investigaciones posteriores. Por ejemplo Lawton et alt. (1997) no encontraron, al igual que Storie, que el sexo del conductor pudiera ser utilizado como un predictor válido del riesgo de accidente, al menos no por si mismo. Sin embargo, lo que plantean Lawton y su equipo es que si se incluye en el estudio de regresión la variable que modela las transgresiones e infracciones, en ese caso el sexo puede ser utilizado como estimador indirecto del

riesgo de accidentalidad, medidado por la mayor tendencia de los hombres a cometer transgresiones de la norma.

Bergdahl y Norris (2001) y, posteriormente Bergdahl (2005) confirman con sus estudios que, a nivel actitudinal los hombres, frente a las mujeres, tienden a cometer mayor número de infracciones por transgresión de normas y muestran menor cumplimiento de la legislación. Estos autores detectan que el hombre parece confiar más en sus habilidades y capacidades como conductor, mientras que la mujer, en general, se muestra más cauta a la hora de evaluar su competencia al volante y es más consciente de posibles limitaciones.

Aún así, gran cantidad de estudios señalan la existencia de diferencias por sexo en cuanto a la conducción como actividad. Lancaster y Ward (2002), en un informe de revisión de las diferencias individuales en la conducta al volante, dan por hecho que existe un mayor riesgo en los hombres de padecer un accidente de tráfico. Anteriormente, Maycock et alt. (1991) habían determinado la existencia de un menor riesgo de accidente en las mujeres para todos los tramos de edad en la que pueden conducir, pero sobre todo en el tramo de jóvenes sin experiencia. Este dato fue confirmado por Lardeli Claret et alt. (2003) en un estudio realizado específicamente en una muestra de la población de conductores españoles en el que también se mostraba la especial relevancia de las diferencias en los grupos de conductores más jóvenes.

Más recientemente Nyberg y Gregesen (2007), ante el hecho de que el número de fallecidos varones en accidente de tráfico es significativamente más alto que el de mujeres, se preocuparon por estudiar las posibles diferencias por sexo que pudieran existir en el proceso de aprendizaje y asimilación de la información necesaria para la conducción cuando se está accediendo a la obtención de un permiso de conducción. Lo que encontraron fue que esta asimilación se da de manera diferente en hombres que en mujeres. De hecho, estas últimas asimilan mejor los contenidos teóricos, realizan una práctica más estructurada, orientada a cubrir posibles carencias y ejecutada en entornos más variados, y, por último, participan más activamente en todo el proceso formativo. Para estos autores, el proceso de asimilación diferencial al que se hace referencia aquí puede ser uno de los factores explicativos que esté mediando en el mayor o menor riesgo de accidente de acuerdo con el género del conductor.

Por último, conviene señalar un aspecto que parece estar relacionado con las diferencias por sexo, pero que no ha sido estudiado en la presente investigación. Sin embargo, tiene suficiente relevancia y significación como para ser motivo de estudio posterior más exhaustivo y detallado en futuros estudios sobre esta temática.

Se trata de los siguiente. Rosenbloom et alt. (2007) centraron su estudio en cómo, durante la realización de un examen práctico de aptitud para acceder a un permiso de conducir, la presencia de un segundo examinado y el sexo de este podría influir en los resultados finales en términos de aprobar o suspender dicha evaluación. El equipo de Rosenbloom encontró que se producían dos fenómenos diferenciados. Por un lado, existía una influencia de la presión social, ejercida no solo por el evaluador sino también por la presencia de un segundo examinado, que se tradujo en que, significativamente, aprobó un mayor número de candidatos de los que realizaron el examen en solitario que de aquellos sujetos que habían realizado el examen acompañados por un segundo examinado.

Por otro lado, dentro de este último grupo, se verificó otro efecto. El número de hombres aprobados fue significativamente superior en el grupo de aquellos que habían realizado el examen acompañados por otro aspirante femenino, frente a aquellos hombres que habían realizado la prueba con la compañía de un segundo examinado varón. Los autores concluyen que parece darse un proceso de facilitación social que media la ejecución de los varones dependiendo del sexo del acompañante.

Este aspecto tan interesante y de tanta relevancia no ha sido estudiado en este proyecto ya que el dato requerido, para informar sobre si la prueba de evaluación se realizó en solitario o en compañía de otros candidatos, no fue recogido y por lo tanto no se registró.

Definitivamente parece suficientemente acreditado, en la literatura científica, la existencia de diferencias entre hombres y mujeres en cuanto a conducción, no solo por lo que se refiere al riesgo de padecer un accidente de tráfico, sino también diferencias en interpretación y cumplimiento de las normas, confianza en las propias capacidades como conductor o asimilación de la información durante el proceso formativo.

Este estudio específico, dentro del contenido general de este trabajo de tesis doctoral, se centra en determinar si la prueba de evaluación práctica que se

presenta aquí es sensible a diferencias motivadas únicamente por el sexo de la persona que realiza la conducción, presentando algún tipo de sesgo, o, si por el contrario, es una prueba que evalúa con la misma precisión a hombres y mujeres sin tener en cuenta el sexo de los candidatos.

Para ello se realizó un estudio de sesgo para verificar la existencia o no de un funcionamiento diferencial de los ítems del test dependiendo del sexo del candidato. Parece razonable y conveniente señalar algunos aspectos relevantes sobre la estimación matemática del funcionamiento diferencial de los ítems, en adelante DIF[11].

5.3.2 Funcionamiento Diferencial de los Ítems

Holland y Thayer (1988) proponen el uso de la expresión Funcionamiento Diferencial de los Ítems (Differential Item Functioning, DIF). Esta denominación cubre aquellos procedimientos estadísticos que pueden ser usados para verificar si un ítem *discrepa* del resto, expresión esta última utilizada a principios de los ochenta por Angoff.

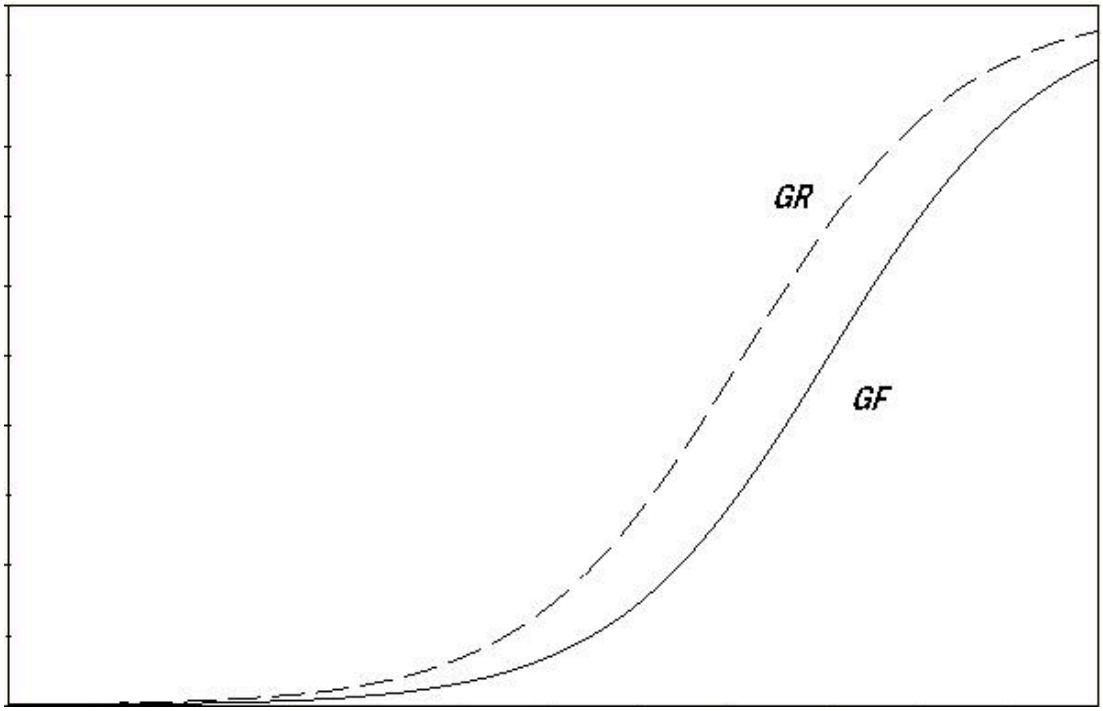
Por lo tanto, puede decirse que un ítem presenta DIF cuando, para un mismo valor de competencia en la variable medida, muestra diferentes propiedades estadísticas en diferentes grupos.

Básicamente, para un mismo nivel de habilidad en la variable medida el ítem no debería mostrar diferencias. Si el ítem reporta diferencias entre los grupos, cuando en la realidad no existen, es cuando se puede hablar de que ese ítem en concreto está discrepando del resto y está mostrando un *funcionamiento diferencial* (DIF).

Mellenberg (1982) distingue dos tipos diferentes de DIF:

1. DIF Uniforme: Es aquel que se da cuando la probabilidad de acierto de un ítem es mayor para un grupo que para otro en todos los niveles de habilidad (Fig. 31).

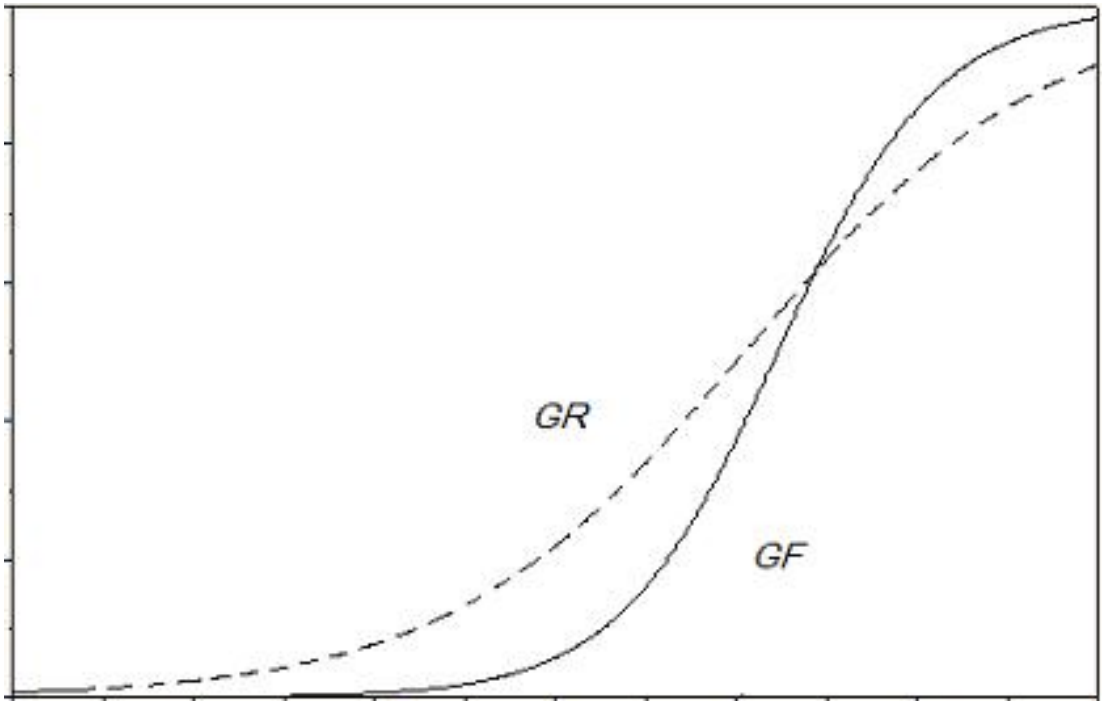
[11] De sus siglas en inglés Differential Item Functioning



Nivel de competencia

Fig. 31: Representación de un ítem con DIF Uniforme

2. DIF No Uniforme: Es el que se da cuando la diferencia entre dos grupos de la probabilidad de contestar correctamente un ítem no es la misma para todos los niveles de habilidad (Fig. 32).



Nivel de competencia

Fig. 32: Representación de un ítem con DIF No Uniforme

Swaminathan y Rogers (1990) proponen el uso del análisis de Regresión Logística para el estudio del DIF. Diferentes estudios han demostrado que se trata de un procedimiento que proporciona resultados muy similares al procedimiento de Mantel-Haenszel cuando existe DIF Uniforme y que tiene una mayor potencia estadística para detectar DIF No Uniforme. Pueden encontrarse referencias a estos trabajos en Hidalgo, Gómez y Padilla (2005).

5.3.3 Estudio y resultados sobre DIF

Los análisis para la realización del estudio de DIF se realizaron sobre la muestra total del estudio general, utilizando únicamente aquellas pruebas de las que se había recogido el dato del sexo de la persona evaluada. El número de pruebas que cumplían ese criterio era de 4335.

Como ya se hizo constar en el apartado de descripción de la muestra (p. 92) la distribución para esta muestra de hombres y mujeres es de 3251 hombres y 522 mujeres. En la figura 33 se representa esta distribución en porcentaje.

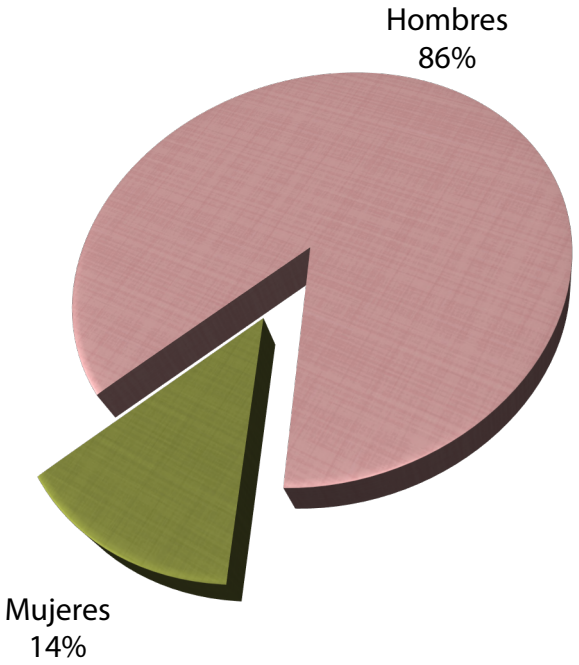


Fig. 33 Distribución de hombres y mujeres (N=4335)

Sobre esta misma muestra, inicialmente, se realizó un estudio de medias para verificar la existencia o no de diferencias significativas en las puntuaciones totales de la evaluación, conforme al sexo del conductor, que pudieran ser indicativas de la existencia de algún tipo de funcionamiento diferencial del test o de alguno de los ítems.

Los resultados de la prueba T de comparación de medias realizada sobre la puntuación del test se presentan a continuación en la Tabla 35.

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias		
		F	Sig.	t	gl	Sig. (bilateral)
Puntuación Total Test	Se han asumido varianzas iguales	**2,295**	**,130**	**14,010**	**3771**	**,000**
	No se han asumido varianzas iguales			**14,734**	**727,089**	**,000**

Tabla 35: Prueba T de comparación de medias para la variable sexo

Los resultados de esta prueba muestran diferencias estadísticamente significativas entre las puntuaciones obtenidas en la prueba por el grupo de hombres frente al grupo de mujeres, siendo la media del grupo masculino superior a la del grupo femenino.

Así las cosas parece sensato profundizar más en el estudio y tratar de resolver la cuestión de si esta diferencia entre ambos grupos se debe a un funcionamiento diferencial del test o de alguno de los ítems presentes en la prueba.

A tal efecto se realizó un estudio del DIF del test utilizando, de entre todos los diferentes métodos posibles, la Regresión Logística. Tal y como se describió anteriormente este método modela la probabilidad de respuesta a un ítem en función de, por un lado, la puntuación empírica obtenida en el test, por otro, de la pertenencia a un grupo y, por último, de la interacción entre ambos factores (Swaminathan y Rogers, 1990).

Incorporando al modelo, de forma sucesiva, los tres factores mencionados se comparan las razones de verosimilitud de los modelos anidados, a saber: [Total]; [Total + Grupo] y [Total + Grupo + Interacción]. Se evalúa el DIF a través del estudio de la mejora del ajuste que se produce en el modelo a medida que se van introduciendo nuevos factores.

En el caso de que la diferencia entre el modelo base y el modelo que incorpora el parámetro de pertenencia al grupo sea estadísticamente significativa se habla de DIF Uniforme. Si esta segunda representación del modelo [Total + Grupo] muestra una diferencia estadísticamente significativa con respecto al modelo completo [Total + Grupo + Interacción] se habla entonces de DIF No Uniforme.

Este tipo de pruebas de significación estadística están fuertemente influenciadas por el tamaño de muestra, siendo común que con muestras grandes se incremente el ratio de errores Tipo I, lo que puede provocar que se determine que un ítem muestra DIF cuando de hecho no es así (Gómez-Benito, Hidalgo, y Zumbo, 2013). Es por esto que, además de este test de significación, en el método de estudio de DIF mediante Regresión Logística se ha propuesto incluir una medida del tamaño del efecto del DIF, lo que proporcionará información más precisa, al basar las decisiones no sólo en su significación estadística sino también en su magnitud.

En este estudio se ha utilizado un método para el cálculo del tamaño del efecto propuesto por Zumbo y Thomas (Thomas y Zumbo, 1996; Zumbo, 1999) basado en las diferencias entre los estadísticos R^2, que representan la proporción de varianza de las puntuaciones en el ítem explicada por el modelo de regresión de los dos modelos (Gierl y McEwen, 1995; Thomas y Zumbo, 1996).

El DIF se considera moderado cuando el estadístico χ^2 es significativo y el incremento de R^2 se sitúa entre 0,035 y 0,070. Tamaños del efecto por debajo de 0,035 implican la no existencia de DIF aún cuando χ^2 sea significativo. Si además de la significación de χ^2, la diferencia entre los dos R^2 es igual o superior a 0,070 el DIF se considera notable (Elosua, 2006).

A continuación se presenta la tabla de resultados tras el estudio del DIF realizado siguiendo el método descrito (Tabla 36).

Ítem	χ^2 (p)	R^2_{M3-M1}	R^2_{M2-M1}	R^2_{M3-M2}
1	3,151 (0,207)	<0,001	<0,001	<0,001
2	0,741 (0,690)	<0,001	<0,001	<0,001
3	2,574 (0,276)	<0,001	<0,001	0
4	6,256 (0,044)	<0,001	0	<0,001
5	53,447 (<0,01)	0,009	0,009	0
6	13,025 (0,001)	0,002	<0,001	0,002
7	21,173 (<0,01)	0,005	0,005	<0,001
8	15,219 (<0,01)	0,003	0,001	0,002
9	18,985 (<0,01)	0,005	0,004	<0,001
10	33,153 (<0,01)	0,005	0,003	0,003
11	17,686 (<0,01)	0,003	0,002	0,001
12	13,851 (0,001)	0,002	0,001	0,001
13	4,267 (0,118)	<0,001	<0,001	<0,001

Tabla 36: Resultados del estudio de DIF

Puede observarse como, en este caso, el tamaño de muestra afecta a la significación estadística y revela que ocho de los trece ítems del test tienen un funcionamiento diferencial. Sin embargo, al combinar la medida del tamaño del efecto puede establecerse que ninguno de los ítems muestra DIF, ya que todos los estadísticos de tamaño del efecto están por debajo del límite establecido de 0,035.

5.4 Atención y resultados en la prueba

5.4.1 Atención y conducción

Existe una importante vinculación entre los problemas atencionales y la accidentalidad en la conducción (Montoro, et al., 2000). Tradicionalmente se han incluido estos problemas dentro del gran cajón de sastre formado por las causas que genéricamente son denominadas *Distracciones*.

La relación viene dada por la relevancia que los procesos atencionales tienen en la percepción. Todo proceso perceptivo comienza por prestar atención a los estímulos que deben ser captados. Esto en un entorno tan rápido, cambiante y tan cargado de estímulos como el tráfico cobra una importancia vital.

En general en la situación de circulación en el tráfico una gran cantidad de los errores perceptivos, errores en la toma de decisiones o incluso fallos de ejecución son causados por una atención inadecuada. Cada vez más la falta de atención adecuada y las diferentes causas de distracción (móvil, gps, múltiple información desde el cuadro de mandos...) son objeto de un estudio más exhaustivo, ya que aparecen detrás de una gran cantidad de los accidentes que se producen en la actualidad.

El consenso es general, las distracciones al volante juegan un papel fundamental como uno de los errores humanos más comúnmente presentes como causa de accidente en carretera (Evans y Schwing, 1985; McKenna, 1982; Montoro, et al., 2000; Van Elslande y Faucher-Alberton, 1996).

Son las capacidades atencionales del conductor las que le permiten dirigir sus recursos mentales hacia determinados aspectos del entorno mientras desecha aquellos otros que no son relevantes para la tarea. De esta forma se favorece la adaptación a la complejidad del entorno del tráfico.

La atención es, por tanto, un proceso clave que actúa de manera relevante para la actividad de la conducción en tres formas principales (Montoro, et al., 2000):

- *Atención Selectiva*: Se trata de un proceso de selección de estímulos clave. Aquellos estímulos relevantes para la tarea son atendidos, mientras que se inhibe la atención de aquellos que no son tan importantes. De esta forma se evita saturar la entrada de información sensorial, optimizando los recursos.

- *Atención Sostenida*: En el caso de la tarea de la conducción, tanto el nivel de alerta (vigilancia) y el tiempo durante el cual este grado de atención puede mantenerse de manera efectiva son clave. El mantenimiento de un adecuado nivel de activación (arousal) mejora sensiblemente la seguridad en la actividad de la conducción.

- *Atención Dividida*: Por último, la capacidad para repartir recursos atencionales entre dos o más fuentes de estímulos, sobre todo cuando estos provienen de diferentes modalidades perceptivas, favorece la automatización de procesos. Esta es una capacidad clave para la conducción, donde gran cantidad de las actividades deben ser automatizadas en aras de, nuevamente, economizar recursos atencionales que serán muy necesarios posteriormente para una adecuada conducción segura.

Indudablemente, la tarea de la conducción requiere el mantenimiento de la atención durante periodos de tiempo muy prolongados, esto produce un aumento de la fatiga lo que, a su vez, genera un peligroso descenso en el grado de activación y alerta, perjudicial para una tarea en la que, en cualquier momento, el conductor deberá estar preparado para percibir un estímulo, evaluar su potencial peligrosidad, tomar una decisión sobre la actuación más adecuada y, por último, llevarla a acabo. Un nivel adecuado de atención, en una situación en la que pueden llegar a recorrerse 33 metros en un segundo, es fundamental para reducir los tiempos de reacción.

Factores externos al conductor como, por ejemplo, una ruta muy familiar, la temperatura del vehículo, las variaciones de luminosidad o una situación del tráfico que capte la atención del conductor afectan a la atención y favorecen las distracciones al volante, pero, de la misma manera, factores propios del conductor como son la fatiga, el sueño, el exceso de estrés o el consumo de alcohol también afectan al proceso atencional, reduciendo su nivel de activación.

Sin embargo, son las llamadas *conductas interferentes* (Montoro, et al., 2007) las que tienen mayor efecto inmediato sobre la atención y, a su vez, son más fácilmente controlables.

Estas conductas interferentes son aquellas que requieren de una atención exclusiva del conductor durante un tiempo lo suficientemente prolongado como para exponerle a un mayor riesgo de accidente. Manejar un mapa, programar el navegador GPS, conversar acaloradamente ya sea con un acompañante o utilizando el manos libres del teléfono, comer o beber mientras se conduce, usar el teléfono móvil, etcétera. Todas son conductas fácilmente evitables, pero con una influencia enorme sobre el descenso de la atención con el consiguiente aumento del riesgo de sufrir un accidente.

En Psicología es habitual el uso de test de evaluación para la medida de la capacidad atencional, bien sea en entornos escolares, clínicos o de selección en departamentos de Recursos Humanos. Uno de los más clásicos y más comúnmente utilizados es el test d2 desarrollado por Rolf Brickenkamp en 1962 y que se desarrolló originalmente como una prueba para evaluar las capacidades atencionales de trabajadores, entre los que se incluían conductores de vehículos pesados (autobuses y camiones) (Brickenkamp, 2009). Es este el test utilizado en este trabajo de tesis para estudiar la capacidad atencional de los candidatos y ponerla en relación con los resultados de la prueba.

5.4.2 Estudio y resultados sobre Atención

Para la realización del estudio se utilizó la prueba d2, descrita anteriormente en el capítulo 3.

Esta prueba de atención se aplicó a una submuestra total de 106 candidatos, obtenida de la muestra general de los participantes que habían acudido a realizar

entrevista de trabajo para su posible incorporación en la empresa. Los evaluados fueron informados de que la prueba que realizaban formaba parte del proceso de selección.

De todos ellos se recogieron los resultados en la evaluación de conducción y los datos de su ejecución en la prueba d2, aplicada siguiendo las instrucciones y procedimientos marcados en la tercera edición de su manual.

En las tablas 40 y 41 se muestran los estadísticos descriptivos de la submuestra, formada por 99 hombres y 7 mujeres. En la primera de las tablas se describen las puntuaciones obtenidas en la prueba de evaluación de la conducción.

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ.	Asimetría
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico
TestTotal	**106**	**10**	**30**	**20,19**	**4,751**	**-,293**
N válido (según lista)	**106**					

Tabla 37: Estadísticos descriptivos Total Conducción (N=106)

En la tabla siguiente se muestran los estadísticos descriptivos de los resultados obtenidos por los candidatos de la submuestra en las diferentes escalas de puntuación de la prueba de atención.

Estadísticos

		D2_TR	D2_TA	D2_O	D2_C	D2_TOT	D2_CON
N	Válidos	**106**	**106**	**106**	**106**	**106**	**106**
	Perdidos	**0**	**0**	**0**	**0**	**0**	**0**
Media		**383,53**	**141,47**	**20,85**	**2,27**	**360,41**	**139,20**
Mediana		**368,50**	**138,00**	**13,00**	**,50**	**350,50**	**136,00**
Desv. típ.		**76,056**	**31,593**	**26,793**	**5,121**	**70,713**	**33,609**
Mínimo		**240**	**60**	**0**	**0**	**226**	**50**
Máximo		**597**	**218**	**146**	**37**	**527**	**217**
Percentiles	25	**320,00**	**120,00**	**7,00**	**,00**	**301,00**	**118,75**
	50	**368,50**	**138,00**	**13,00**	**,50**	**350,50**	**136,00**
	75	**439,50**	**161,00**	**24,00**	**2,00**	**419,25**	**158,75**

Tabla 38: Estadísticos descriptivos Prueba de Atención d2 (N=106)

Análisis de Varianza

A partir de aquí se llevó a cabo un Análisis de Varianza (ANOVA) para verificar la posible existencia de diferencias en la puntuación final de la evaluación de conducción en el test total en función de los diferentes resultados obtenidos en las escalas de la prueba d2. Esto es, verificar si mejores niveles atencionales mejoran el rendimiento global en la prueba.

A continuación se muestran los resultados del Análisis de Varianza realizado a la submuestra que se describió anteriormente (Tabla 41). En este caso se realizó el estudio tomando la escala TR del d2 como variable independiente. Tal y como se mencionó esta escala se define por el Total de Respuestas emitidas y se considera un indicador fiable de la atención (sostenida y selectiva) del sujeto que responde al test. Para llevar a cabo el análisis se agruparon las puntuaciones en esta escala formando tres grupos. Se asumió que un TR bajo agrupará a puntuaciones inferiores a 320 puntos y que un TR alto estará formado por las puntuaciones superiores a 439 puntos, de acuerdo a la distribución de puntuaciones obtenidas en la muestra estudiada.

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
TestTotal	Inter-grupos	**15,281**	**2**	**7,640**	**,334**	**,717**
	Intra-grupos	**2354,946**	**103**	**22,864**		
	Total	**2370,226**	**105**			

Tabla 39: Análisis de varianza para Puntuación en el Test por nivel atencional TR

Puede comprobarse que el análisis no muestra la existencia de diferencias estadísticamente significativas en las puntuaciones obtenidas en la evaluación de conducción teniendo en cuenta el nivel de atención medido por el nivel del Total de Respuestas dadas en el test d2.

Aunque el gráfico de medias marginales (Fig. 34) muestra una tendencia hacia la obtención de medias más altas en la prueba de conducción cuanto mayor es el nivel de TR medido, estas diferencias no tiene entidad suficiente como para ser estadísticamente significativas.

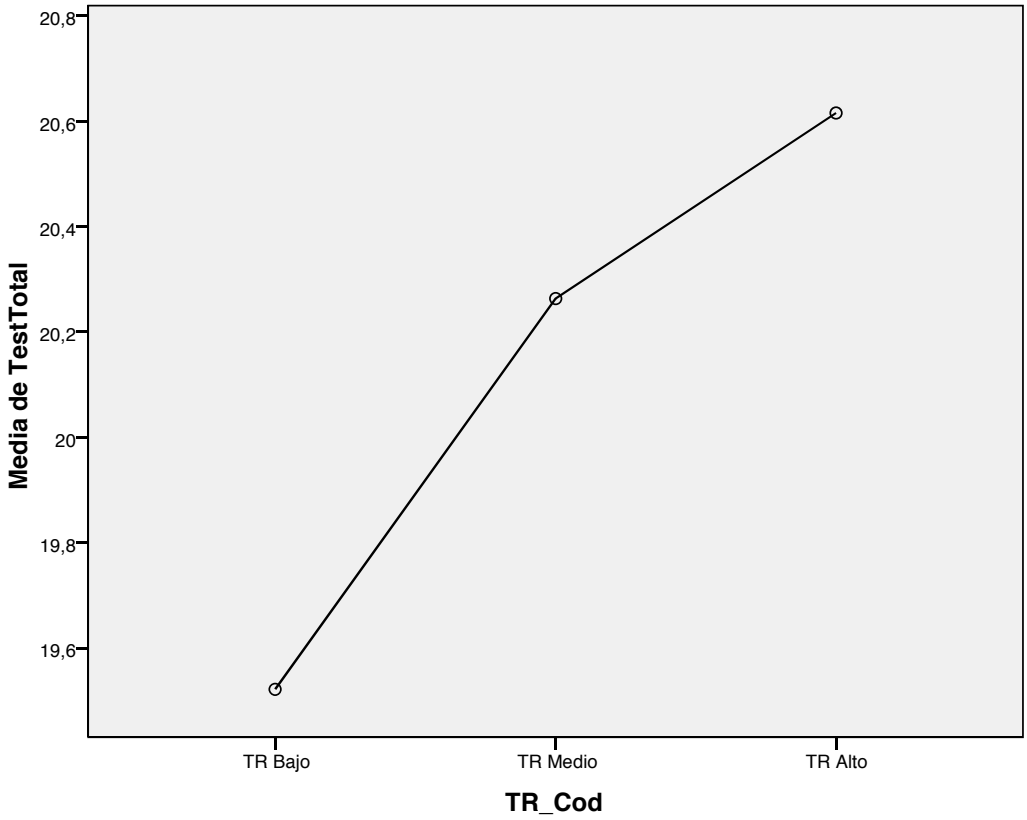


Fig. 34 Gráfico de medias marginales estimadas ANOVA nivel TR

Los resultados obtenidos en los análisis de varianza realizados utilizando las demás escalas del test d2 arrojan resultados similares a los presentados para la escala TR. A modo de información resumen se muestran, en la siguiente tabla, los valores del estadístico F obtenidos y su probabilidad asociada (Tabla 40). En el anexo final de esta memoria están recogidos los resultados completos de cada uno de los diferentes análisis.

Escala d2	F	Sig.
TR (Total de Respuestas emitidas)	0,334	0,717
TA (Total respuestas Acertadas)	0,224	0,800
O (Omisiones, respuestas no marcadas)	0,376	0,688
C (Comisiones, marcas incorrectas)	0,073	0,930
TOT (Puntuación total)	0,938	0,395
CON (Nivel de concentración)	0,375	0,688

Tabla 40: Resumen de ANOVA para escalas d2

Análisis de tablas de contingencia

Al observarse, en algunas de las escalas de la prueba de atención, cierta tendencia que podría relacionar el nivel de atención con los resultados obtenidos por los sujetos de la muestra en la evaluación de conducción y a pesar de no existir diferencias estadísticamente significativas, se llevó a cabo un estudio utilizando tablas de contingencia. Para ello se utilizó la misma muestra que había completado el test de atención (N=106) con el objetivo de ver si un análisis desde otro enfoque permitía identificar relaciones entre las variables atencionales y la ejecución en la prueba de conducción.

El análisis utilizando tablas de contingencia sirve para estudiar la relación entre dos o más variables entre las que , generalmente al menos una de ellas es categórica o cualitativa. La técnica distribuye el número de observaciones entre todas las combinaciones posibles de las variables presentes en el estudio.

En la Tabla 41 se muestra, como representación, la tabla de contingencia que se calculó para poner en relación la puntuación obtenida en la prueba de conducción (filas) con el nivel categórico (Bajo, Medio y Alto) de la puntuación en Total Respuestas en el test d2.

Tabla de contingencia TestTotal * TR_Cod

Recuento

		TR_Cod			
		1	2	3	Total
TestTotal	10	**1**	**1**	**0**	**2**
	11	**0**	**2**	**0**	**2**
	12	**1**	**3**	**1**	**5**
	13	**0**	**1**	**1**	**2**
	14	**0**	**3**	**2**	**5**
	15	**0**	**3**	**1**	**4**
	16	**2**	**0**	**1**	**3**
	17	**4**	**5**	**0**	**9**
	18	**3**	**2**	**5**	**10**
	19	**1**	**1**	**1**	**3**
	20	**3**	**2**	**2**	**7**
	21	**2**	**1**	**0**	**3**
	22	**3**	**2**	**0**	**5**
	23	**2**	**13**	**3**	**18**
	24	**1**	**4**	**3**	**8**
	25	**2**	**5**	**2**	**9**
	26	**1**	**3**	**1**	**5**
	27	**1**	**0**	**1**	**2**
	28	**0**	**1**	**0**	**1**
	29	**0**	**1**	**1**	**2**
	30	**0**	**0**	**1**	**1**
Total		**27**	**53**	**26**	**106**

Tabla 41: Tabla de contingencia Puntuación total en test por total respuestas en el test de atención

Si la proporción de casos en cada columna varía entre las filas se debe considerar la posibilidad de no independencia entre las variables estudiadas (Agresti y Finlay, 2009).

Para poner a prueba esta variación cuando se trabaja con tablas de contingencia se utiliza el estadístico χ^2 de Pearson el cual contrasta las siguientes hipótesis:

$$H_0 = \textit{variables estadísticamente independientes}$$

$$H_1 = \textit{variables estadísticamente dependientes}$$

Para ello el estadístico compara las frecuencias observadas en la tabla de contingencia con las frecuencias esperadas que satisfagan la hipótesis nula. La formulación es la siguiente:

$$\chi^2 = \sum_{1}^{l} \frac{(f_o - f_e)}{f_e}$$

La aceptación de la hipótesis nula permite considerar a las variables implicadas como estadísticamente independientes.

En la Tabla 42 se muestran los resultados obtenidos del cálculo del χ^2 de Pearson sobre los datos de la tabla anterior.

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	**39,855**[a]	**40**	**,477**
Razón de verosimilitudes	**48,553**	**40**	**,166**
Asociación lineal por lineal	**,341**	**1**	**,559**
N de casos válidos	**106**		

Tabla 42: χ^2 de Pearson Puntuación total en test por total respuestas en el test de atención

Además del estudio del estadístico χ^2 de Pearson que indica la existencia o no de independencia entre variables, se puede calcular una serie de *medidas de asociación* que permiten cuantificar la relación existente entre los diferentes criterios de clasificación (Álvarez, 2000).

Existen diversos coeficientes para la medida de esa asociación entre los distintos atributos de la tabla de contingencia. En el caso que aquí se trata, con una escala de intervalo y otra escala de tipo ordinal, el coeficiente recomendado como medida de asociación es el coeficiente Eta. Valores de Eta cercanos a 1 son indicativo de una fuerte dependencia entre variables, de la misma manera, valores próximos a 0 se relacionan con variables independientes, o al menos con un grado bajo de relación.

La Tabla 43 recoge el valor de Eta para la tabla de contingencia que relaciona la puntuación total en el test de conducción y el total de respuestas emitidas en el test de atención.

Medidas direccionales

			Valor
Nominal por intervalo	Eta	TestTotal dependiente	**,058**
		TR_Cod dependiente	**,432**

Tabla 43: Coeficiente Eta Puntuación total en test por total respuestas en el test de atención

Como puede observarse, ni el estudio de hipótesis mediante χ^2 de Pearson ni los valores de Eta obtenidos permiten pensar en la existencia de relación alguna entre la puntuación que los participantes obtienen en la prueba de evaluación de conducción y la escala TR del test de atención d2.

Los resultados obtenidos en las tablas de contingencia realizadas con el resto de las escalas del test de atención son similares a los mostrados para la escala TR.

De nuevo se presenta una tabla resumen con los valores más relevantes obtenidos, en este caso los valores de χ^2 de Pearson y de Eta para cada escala del test de atención relacionada mediante una tabla de contingencia con la puntuación total del test de evaluación de conducción (Tabla 44). En el anexo correspondiente a este capítulo, se presentan los resultados completos de todos los estudios.

Escala d2	Sig χ^2	Eta
TR (Total de Respuestas emitidas)	0,477	0,432
TA (Total respuestas Acertadas)	0,525	0,327
O (Omisiones, respuestas no marcadas)	0,772	0,398
C (Comisiones, marcas in correctas)	0,889	0,396
TOT (Puntuación total)	0,563	0,384
CON (Nivel de concentración)	0,481	0,372

Tabla 44: Resumen valores χ^2 de Pearson y Eta para tablas de contingencia

5.5 Personalidad y resultados de la prueba

5.5.1 Personalidad y conducción

Si existe un tópico ampliamente relacionado con la conducción y la accidentalidad en carretera ese es, sin ningún género de duda, el de la personalidad.

Gran parte de la literatura relacionada con este aspecto señala la existencia de algún tipo de relación, en diferente grado, entre distintos perfiles de personalidad y la conducción de un vehículo. En la mayoría de las ocasiones esta relación se da tomando los factores de personalidad como elemento mediador relacionado directamente con aspectos propios de lo que se ha denominado conducción de riesgo (Kalyoncouglu y Tigdemir, 2008; Schwebel, et al., 2007).

La excelente revisión realizada por Jonah (1997) sobre cuarenta estudios diferentes, en los que de una forma u otra se evaluaba la conducción de riesgo, o alguno de sus factores concretos, aporta una visión bastante clara de dónde y en qué grado se da la relación entre personalidad y conducción.

En general, la relación más fuerte se encuentra entre la *Búsqueda de sensaciones,* como rasgo de la personalidad, y algunos aspectos que se encuentran relacionados directamente con la conducción de riesgo o conductas de riesgo durante la conducción. Por ejemplo, en siete de los estudios revisados por el autor aparecen diferencias estadísticamente significativas entre niveles altos de necesidad de búsqueda de sensaciones y la implicación en un mayor número de accidentes de

tráfico. Una descripción pormenorizada de este estudio puede verse en Lancaster y Ward (2002).

Las conclusiones de Jonah se sitúan en la línea de considerar, a la luz de los resultados de las investigaciones revisadas, que las personas con niveles altos del rasgo *búsqueda de sensaciones* tienden a percibir como menos peligrosas situaciones propias de la conducción que, de manera general, se aceptan como situaciones de riesgo real de sufrir un accidente de tráfico. Esta menor percepción del riesgo se operativiza en una mayor probabilidad de padecer un accidente o de ser multados por algún tipo de infracción de tráfico.

De la misma forma en que esta relación entre la necesidad de búsqueda de sensaciones y la conducción de riesgo aparece bastante clara en la literatura y en diferentes investigaciones relacionadas a lo largo de los últimos años (Iversen y Rundmo, 2002; Oltedal y Rundmo, 2006), no es tan sencillo encontrar una relación tan clara entre otros perfiles de personalidad y accidentalidad en carretera. Esta relación, si existe, varía y parece mezclarse entre los diferentes factores de personalidad (Lancaster y Ward, 2002). Tal vez el aspecto más estudiado en relación con la personalidad y la conducción han sido las escalas del cuestionario de personalidad de Eysenck, especialmente en los continuos extraversión-introversión y neuroticismo-control.

De forma general, el tipo de personalidad extravertido presenta una serie de características que pueden facilitar la aparición de patrones de conducta que se asocian con estilos de conducción considerados de riesgo (Montoro, et al., 2000). Su mayor impulsividad, pérdida de control personal, baja condicionabilidad, la poca precisión y exigencia en sus tareas o su baja inhibición la cual convierte al tipo extravertido en un buscador de emociones, son características que colocan a este perfil de personalidad en situación de riesgo ante la posibilidad de sufrir un accidente de circulación.

De la misma manera, la capacidad de concentración de su atención, su facilidad para condicionar o el nivel alto de cautela, obediencia y rechazo por las emociones fuertes, protegen al introvertido frente a las situaciones de riesgo que se pueden dar en la conducción.

Por lo que se refiere al grado de control emocional de la conducta, el tipo de personalidad neurótico se caracteriza por un mayor nivel de actividad emocional,

alto nivel de impulsividad, bajo autoconcepto y, en general, una puntuación inferior al promedio en agudeza sensorial. Estas son muchas de las características que aparecen con frecuencia en sujetos con una tasa alta de accidentes de tráfico (Montoro, et al., 2000).

Parece claro, en opinión de los autores citados, que el conductor ideal se sitúa en el polo *introvertido-controlado* de ambos continuos. Estudios como el de Shaw y Sichel (1971) citado en Furnham y Saipe (1993) apoyan esta tesis. En su estudio, realizado con conductores de autobús, los autores determinaron que aquellos conductores que presentaban peor registro de accidentes se podrían clasificar dentro del polo *extravertido-neurótico,* mientras que los que tenían mejor registro en cuanto a accidentalidad se refiere eran de tipo *introvertido*.

Las investigaciones sobre personalidad y tráfico han venido utilizando una gran variedad de los test y cuestionarios que existen a disposición para evaluar las diferentes facetas de la personalidad. Como se ha visto el cuestionario de personalidad EPI ha sido y es uno de los más utilizados. No obstante, también se han llevado a cabo estudios utilizando otras herramientas conocidas y frecuentes en el estudio de la personalidad. Este es el caso, por ejemplo, del Cuestionario Factorial de Personalidad de Cattell, 16PF. En un estudio clásico Hilakivi y otros (1989) estudiaron el valor predictivo de este test en la ocurrencia de accidentes de tráfico en una muestra de jóvenes reclutas durante su servicio militar, realizado en la sección de logística y transporte del ejército. Tras realizar la investigación sobre una muestra de 597 sujetos, de los que se había registrado su experiencia anterior, su número de accidentes e historial de multas, los autores concluyeron que los mejores predictores de accidentalidad de entre todos los factores presentes en el 16PF eran:

- Elevada impulsividad. Factor H alto.
- Exceso de confianza. Factor L bajo.
- Escaso autocontrol. Factor Q_3 bajo.

Estos tres factores citados por los autores del estudio, son los que predicen significativamente la ocurrencia de accidentes.

En la actualidad, ya desde principios de los años ’90, ha ido cobrando fuerza un modelo de representación de la estructura de los rasgos de personalidad, elaborado por diversos autores, entre los que se encuentran Digman, McCrae y

Costa. Este modelo (Digman, 1990) representa en cinco factores las dimensiones de la personalidad que subyacen a su descripción, tanto en el lenguaje cotidiano (John, 1990) como en la mayoría de los test y cuestionarios psicológicos de evaluación de la personalidad (McCrae y Costa, 1987, 1990).

En este modelo, cuando se describe la puntuación de un sujeto en cada uno de los cinco factores, se obtiene un perfil de su estilo de personalidad a nivel emocional, interpersonal, experiencial, actitudinal y motivacional.

Estos cinco factores, conocidos como *Big Five* (Ewen, 1998), proporcionan un conocimiento básico de la personalidad del sujeto. Un análisis más detallado de las diferentes facetas de cada uno de esos factores aporta un análisis más detallado y preciso de los rasgos específicos dentro de cada uno de ellos.

5.5.2 Estudio y resultados sobre personalidad

Para la realización del estudio específico sobre los posibles efectos de la personalidad en las puntuaciones de la prueba práctica de evaluación de conducción se tomó una submuestra de los participantes en los procesos de selección, durante los cuales utiliza la prueba como herramienta de evaluación. A estos participantes se les incluyó, dentro del conjunto total del proceso, la aplicación de la forma reducida del test NEO PI-R descrito en el capítulo 3.

La submuestra está formada por 101 sujetos de los que se recogió, en la realización de su proceso de selección, su puntuación en la evaluación práctica de conducción, utilizando el test aquí estudiado,junto con los resultados de la realización del test NEO PI-R.

En la Tabla 45 se muestran las puntuaciones medias en las cinco facetas del test y la distribución de los cuartiles.

Estadísticos

		NEO Neuroticismo	NEO Extraversión	NEO Apertura	NEO Amabilidad	NEO Responsabilidad
N	Válidos	**101**	**101**	**101**	**101**	**101**
	Perdidos	**3756**	**3756**	**3756**	**3756**	**3756**
Media		**13,81**	**33,42**	**25,56**	**34,17**	**37,71**
Percentiles	25	**10,00**	**28,50**	**22,00**	**32,00**	**34,00**
	50	**12,00**	**33,00**	**25,00**	**34,00**	**39,00**
	75	**18,50**	**39,00**	**28,00**	**38,00**	**42,00**

Tabla 45: Estadísticos descriptivos NEO PI-R

El estudio se realizó utilizando la técnica de Análisis de Varianza univariante recogida en el paquete estadístico SPSS v.17. Para ello, previamente, se agruparon las puntuaciones en el test de personalidad en tres grupos diferentes definidos como Nivel Bajo (1), Nivel Medio (2) y Nivel Alto (3).

La Tabla 46 muestra esa agrupación para la variable Neuroticismo (NEO_N). Este proceso se repitió para las otras cuatro variables de personalidad, sus tablas de distribución de frecuencias se pueden ver en el anexo final.

NEO_N_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**32**	**,8**	**31,7**	**31,7**
	2,00	**44**	**1,1**	**43,6**	**75,2**
	3,00	**25**	**,6**	**24,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

Tabla 46: Distribución de puntuaciones agrupadas Neuroticismo NEO PI-R

A continuación en la Tabla 47 se presentan los resultados del Análisis de Varianza realizado. En esta tabla se incluyen únicamente los efectos del modelo completo junto con los efectos principales de las variables de personalidad tomadas de forma independiente, sin tener en cuenta las interacciones entre ellas y su efecto sobre la variable dependiente.

En el anexo se incluye la tabla del Análisis de varianza completo.

Pruebas de los efectos inter-sujetos

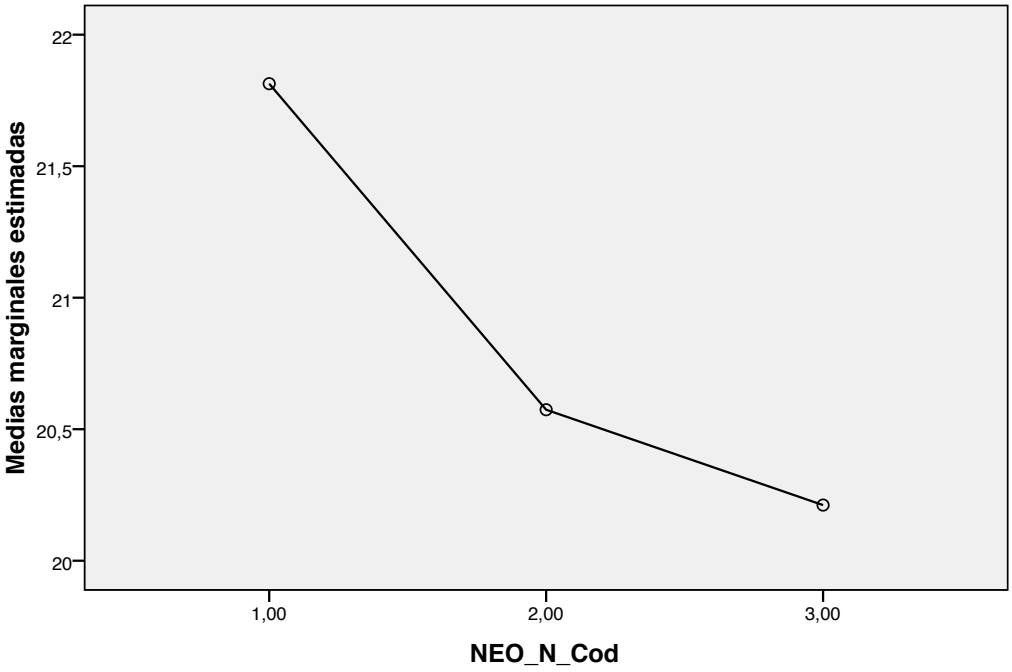
Variable dependiente:Puntuación Total Test

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	**1092,367**[a]	**47**	**23,242**	**,674**	**,914**
Intersección	**22992,372**	**1**	**22992,372**	**667,145**	**,000**
NEO_N_Cod	**14,506**	**2**	**7,253**	**,210**	**,811**
NEO_E_Cod	**15,840**	**2**	**7,920**	**,230**	**,795**
NEO_O_Cod	**54,860**	**2**	**27,430**	**,796**	**,456**
NEO_A_Cod	**73,051**	**2**	**36,525**	**1,060**	**,354**
NEO_C_Cod	**11,227**	**2**	**5,613**	**,163**	**,850**

Tabla 47: Análisis de varianza para puntuación en el test por resultados en el test NEO PI-R de personalidad

Puede observarse que ni el modelo completo ni los factores de personalidad tomados de forma individual muestran diferencias significativas en el análisis.

Por último, de los gráficos de medias marginales estimadas únicamente los de la variable Neuroticismo y la variable Responsabilidad muestran una tendencia claramente identificable, pero poco significativa. En el caso de la primera de ellas muestra tendencia a obtener puntuaciones más bajas a medida que aumenta el nivel de Neuroticismo. En el segundo de los casos la tendencia es inversa, mayores puntuaciones a medida que aumenta el nivel de Responsabilidad (Fig. 35).



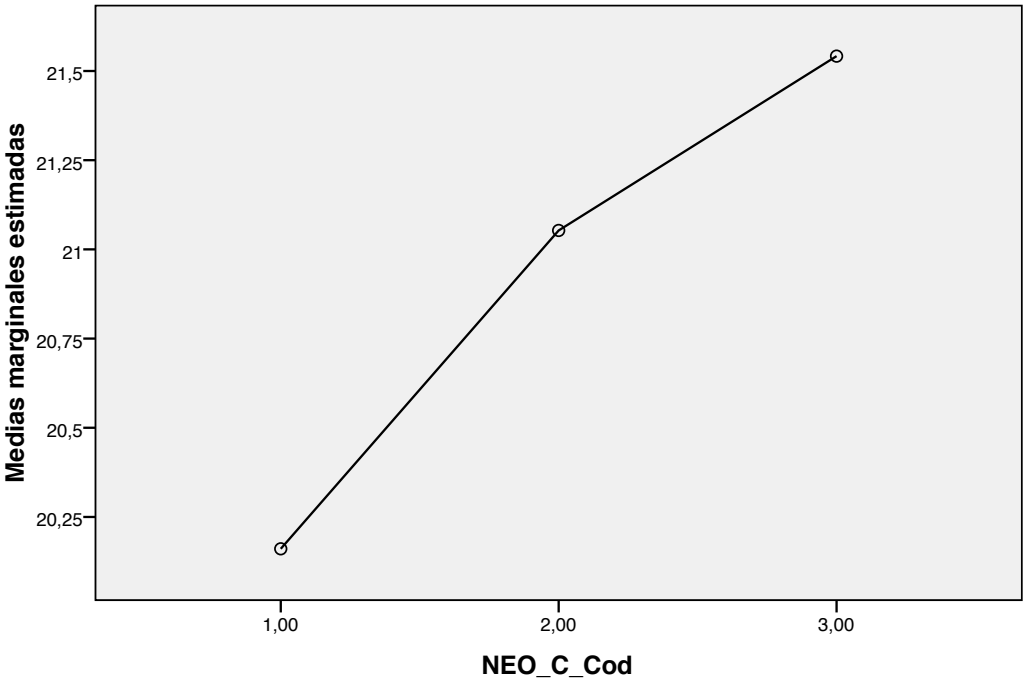


Fig. 35: Gráficos de medias marginales estimadas, Neuroticismo (N) y Responsabilidad (C)

Un científico debe tomarse la libertad de plantear cualquier cuestión, de dudar de cualquier afirmación, de corregir errores.

Julius Robert Oppenheimer (1904-1967)

Discusión de los Resultados 6

En el presente capítulo se profundiza en los resultados de los estudios presentados. Se hace de una manera estructurada, respetando lo más posible el orden en que se presentaron los distintos estudios en el texto.

En una primera sección se discuten los resultados descritos, relacionados con las propiedades psicométricas de la prueba, presentado información sobre la Validez, Fiabilidad y el Análisis de ítems tanto desde la perspectiva de la Teoría Clásica de los Test como desde la Teoría de Respuesta al Ítem.

En la segunda sección de este capítulo se discuten, punto por punto, los resultados de los cinco estudios específicos realizados en la presente investigación como complemento al estudio de las propiedades de medida de la prueba. Al igual que en la primera sección, la discusión de estos resultados se presenta en el orden en que se describieron los estudios en el capítulo precedente.

6.1 Discusión de las Propiedades Psicométricas

El objetivo principal de esta tesis doctoral gira en torno al estudio de las propiedades psicométricas, las propiedades de medida, de una prueba práctica para la evaluación de la habilidad en la conducción específicamente de conductores de autobús. Tal y como se reflejó en el tercer capítulo de este texto, al describir el proceso de construcción esta herramienta, la elaboración de este test no respondió estrictamente a un proceso *psicométrico* en sentido puro y controlado. Con esto se pretende manifestar que la prueba nace inicialmente como una herramienta de ayuda, con unos objetivos concretos de evaluación en los procesos de selección, y en el proceso de su construcción únicamente la experiencia y el sentido común de quienes la desarrollaron, que no tenían ni conocimiento ni relación con el campo de la Psicometría, fueron los instrumentos para el desarrollo de la prueba inicial con la que se empezó a trabajar.

Los pasos dados posteriormente en su revisión y mejora, motivados por la intención de su propietario de que esta prueba se convierta en el eje central de todo un proceso de evaluación y formación que se ha puesto en marcha en su Organización, han tratado de ir más acordes con el proceso estructurado de construcción de una herramienta de este tipo, ajustándose en la medida de lo posible a los cánones psicométricos. Todo esto sin perder de vista la necesidad de su utilidad práctica inmediata y los plazos ajustados marcados principalmente por el ritmo de producción empresarial, traducido en este caso en la necesidad de asignar nuevos conductores a los servicios en unas condiciones garantizadas de rendimiento y seguridad.

Fruto de ese esfuerzo por reformar la prueba original, adaptarla y hacerla más fácil de utilizar y aplicar se diseña la prueba que actualmente se utiliza en todos los procesos de este tipo que se dan en la empresa. Aquí ha sido utilizada para investigar sus propiedades de medida como herramienta en uso y sobre la que se basan algunas de las decisiones que se toman con respecto a la valía de los candidatos y conductores que la realizan. Desde esa perspectiva, un análisis profundo de sus propiedades de medida proporcionará la base definitiva para considerar esta prueba como una herramienta fiable y válida para la tarea para la que fue diseñada y pensada, reforzando así el peso de las decisiones que se tomen a partir de los resultados de su uso.

6.1.1 Evidencias de Validez

Como ya se manifestó en capítulos anteriores la perspectiva psicométrica actual contempla la Validez como un concepto unitario para el que existen diferentes estrategias de aproximación. La Validez de Contenido, la Validez de Constructo o la Validez Predictiva son diferentes formas de evaluar la capacidad que un instrumento tiene para medir aquello para lo que fue construido y no otra cosa distinta.

Siguiendo la recomendación de la APA (1999) de incluir, en el proceso de validación de cualquier instrumento de medida en psicología, evidencias de validez obtenidas de diferentes formas, utilizando diferentes fuentes y modos de validez para el presente trabajo se realizó el estudio de Validez desde el punto de vista de la estrategia de desarrollo y construcción del contenido del test, se evaluó la Validez de constructo a partir de un análisis de la validez factorial del test y, por último, se realizó un estudio, complementariamente al estudio de Validez, para valorar la Sensibilidad de la prueba utilizando una prueba de Regresión Logística al carecer datos en condiciones adecuadas para un análisis más profundo de la Validez Predictiva de la prueba.

Evidencias basadas en el contenido del test

Con respecto a esta validación del test, como se ha señalado en ocasiones anteriores, debe tenerse en cuenta que el origen y la creación de esta herramienta no son ortodoxos, ni han seguido la doctrina *Psicométrica* al uso para este tipo de trabajo de construcción de una herramienta de medida. La prueba original no fue construida por expertos psicómetras ni validada posteriormente siguiendo algún procedimiento estandarizado.

Si que sirvió como punto de partida para todo el trabajo posterior que se realizó hasta llegar al desarrollo de la actual prueba que se está validando con la elaboración de esta tesis doctoral.

Las modificaciones que se plantearon sobre la prueba inicial fueron acometidas a partir del trabajo de un panel de expertos orientados por psicólogos conocedores de la materia y del procedimiento. Se eliminaron aquellos ítems de la prueba inicial de los cuales el grupo de expertos consideró que, conceptualmente, no se relacionaba con el constructo propuesto inicialmente, *Habilidad en la conducción*, y definido tal cual se ha señalado anteriormente:

> Capacidad de actuar que se desarrolla gracias al aprendizaje, al ejercicio y a la experiencia y que, para el caso de la conducción de vehículos, implica la ejecución de una serie de técnicas y la puesta en marcha de un esquema de conducta relacionado con la Seguridad y Eficiencia Energética.

Así, por ejemplo, se eliminaron de la batería de ítems del test inicial aquellos que hacían referencia a las nociones de uso del Tacógrafo y cumplimentación de su correspondiente disco diagrama, ya que parecía evidente que no aportaba nada a la habilidad en la conducción, así descrita, y perjudicaba sistemáticamente a los conductores menos experimentados y a los que, aún teniendo experiencia, esta se circunscribía a vehículos pesados exentos de la obligatoriedad de uso del tacógrafo analógico.

El panel agrupó los ítems siguiendo el esquema de la definición del constructo creando tres secciones dentro del test.

En primer lugar aquellos ítems considerados relevantes para la medida de las conductas relacionadas con la eficiencia energética. Estos incluyen desde el correcto control del funcionamiento del régimen motor, manteniendo las revoluciones dentro de la denominada *zona económica del par*, hasta la gestión y anticipación de situaciones del tráfico, de tal forma que la forma de circular favorezca un menor consumo anticipando cambios en el perfil de la carretera, aprovechando las inercias propias del vehículo y anticipando frenadas. El uso racional del mecanismo retardador de la frenada y la decisión sobre el momento correcto de uso del freno de servicio complementan esta sección.

En el segundo apartado se agruparon los ítems que evalúan el esquema de conducta relacionado con la seguridad. En esto el control emocional es el punto inicial.

A partir de aquí se le exige al candidato seguridad y firmeza en el manejo del volante y tres actitudes que el panel de expertos consideró básicas para la seguridad vial, la observación constante de la información proporcionada por los retrovisores del vehículo, la permanente atención y respeto a la señalización vial del recorrido y, por último, la circulación en el vehículo manteniendo una distancia mínima de seguridad con el vehículo precedente, tanto al circular como al detenerse.

En la sección final se evalúan las habilidades entendidas en el sentido más estricto, por lo que recoge aptitudes de conducción técnica. Estas implican un correcto uso del cambio de marchas, la permanente observación de las indicaciones del cuadro de mandos, la capacidad para integrarse en el tráfico adaptándose a sus condiciones cambiantes circulando de una forma fluida y, por último, un necesario sentido de la maniobrabilidad que demuestre conciencia de las dimensiones del vehículo y su capacidad de giro para realizar todo tipo de maniobras, desde las más sencillas como aquellas más complejas programadas por el evaluador durante el recorrido.

Una parte muy importante del trabajo del panel de expertos se centró fundamentalmente en la definición clara de las realizaciones conductuales que permitían determinar el nivel del candidato en cada uno de los ítems. La descripción que se hace en el capítulo 3 de este texto de cada uno de los ítems refleja el esfuerzo depositado en, no solo definir claramente cada uno de los niveles, sino también en conseguir un consenso general de tal forma que se evitara, en la medida de lo posible, la interpretación libre de cada evaluador a la hora de juzgar una puntuación u otra.

Con lo aquí expuesto puede decirse que el trabajo que se realizó en la construcción de la herramienta ha tratado de salvar, en la medida de lo posible, las limitaciones iniciales producidas por la herencia de una prueba previa *no psicométrica* y recoger de una forma organizada, sistemática y definida todos los aspectos relacionados con la definición del constructo a evaluar.

Con esto puede decirse que la Validez de Contenido de la prueba queda cubierta al verificarse que, en la elaboración de la prueba, se ha cubierto una muestra representativa de todos aquellos aspectos relevantes para el objeto de estudio

Evidencias basadas en la estructura interna del test

Al inicio de la investigación se hipotetizó que el test no solo debería medir la habilidad en la conducción, definida como constructo general, sino que, si el constructo se había definido correctamente, a pesar de tomarse tres aspectos de

manera diferenciada para la evaluación de la conducción (Economía, Seguridad y Técnica), la estructura subyacente del test debería ser unidimensional, dando el test, de esta manera, información de un único factor general que será considerado, conforme a la definición realizada del constructo, como *Habilidad en la conducción*.

Para evaluar esta premisa se realizó un estudio de la Validez de constructo mediante la técnica de Análisis Factorial. Siguiendo la recomendación propuesta por Costello y Osborne (2005) el Análisis Factorial no se limitó únicamente a un estudio exploratorio (AFE) sino que incluyó también un estudio confirmatorio (AFC) de la estructura obtenida en el primer análisis. De esta manera, como ya se explicó, el análisis exploratorio se realizó utilizando una submuestra obtenida de la muestra general del estudio. Posteriormente el análisis confirmatorio utilizó una segunda submuestra de la muestra general, distinta de la primera, para confirmar la estructura obtenida en el AFE. Este proceso de cruce de técnicas de validación aporta mayor credibilidad a los datos del ajuste del modelo confirmatorio ya que utiliza una muestra, tomada de la misma población, pero distinta de la utilizada para determinar la estructura subyacente al modelo en el proceso exploratorio.

Por congruencia con el proceso de construcción del test, anteriormente descrito, y por existir la posibilidad de que la estructura se revelara de esa forma, inicialmente, en el análisis exploratorio se programó una estructura inicial de 3 factores, a pesar de que lo esperado era que la estructura más explicativa fuera una estructura unidimensional.

En los resultados del AFE presentados en esta investigación puede comprobarse que el resultado más parsimonioso y que ofrece una mejor explicación de los datos es la interpretación de que la estructura es unidimensional. Todos los ítems presentan valores de saturación por encima de 0,30 en el Factor 1.Por otro lado, las saturaciones en los otros dos factores extraídos inicialmente son bajas, en general, para todos los ítems y no permiten pensar en la existencia de un segundo factor subyacente.

Tomado el criterio clásico de autovalores mayores que 1 podría considerarse la existencia de un segundo factor, pero la escasa diferencia que separa al segundo autovalor del tercero (0’13655) se reflejaría en una gráfica scree-plot asintótica desde el segundo autovalor, lo cual es más congruente con la interpretación de una estructura unidimensional. Además, la introducción de un segundo factor no mejora la explicación inicial ni aporta más claridad a la estructura.

Es también valorable el hecho de que las comunalidades de las variables, porcentaje de varianza del ítem explicada por el factor común, son bastante altas en general y especialmente en el ítem número 13 *Sentido de la maniobrabilidad*, el cual es, inicialmente, el más similar a la descripción general de habilidad en la conducción, ya que es el que exige un mayor compendio de habilidades puestas en marcha a la hora de conducir el vehículo.

Finalmente la correlación entre factores es alta; 0,624 entre el Factor 1 y el Factor 2; 0,608 entre el Factor 1 y el 3; 0,580 entre el 2 y el Factor 3. Esto es indicativo de una estructura más propia de un único factor ya que los tres se muestran muy relacionados entre si.

El análisis factorial de segundo orden que se realizó a partir de las puntuaciones factoriales, obtenidas en el primer estudio exploratorio, arroja también como resultado una estructura unifactorial.

Asumiendo esta estructura de un único factor como la más plausible, la más parsimoniosa y la que mejor explicaba el modelo inicialmente hipotetizado, se realizó un Análisis Factorial Confirmatorio para verificar el ajuste de este modelo unidimensional.

Los datos que revelan los índices de ajuste obtenidos son muy buenos en prácticamente todos los índices y son congruentes con la hipótesis de una estructura unidimensional subyacente en el test. El ajuste no es bueno únicamente en los dos índices de ajuste basados en χ^2 (CMIN y CMIN/df). Como también se ha señalado en otros apartados de este estudio χ^2 es prácticamente siempre significativo en muestras superiores a 200 sujetos, lo cual explicaría el mal dato de ajuste que aportan ambos índices. Como está recomendado se estimaron los índices de Hoelter para alpha 0,05 y alpha 0,01, para verificar el límite superior del tamaño muestral a partir del cual la indicación de ajuste de χ^2 deja de ser fiable. En este caso ambos datos, Hoelter = 435 sujetos para un nivel de significación alpha 0,05 y Hoelter = 508 sujetos para un nivel de significación alpha 0,01, son límites muy por debajo del tamaño muestral utilizado para el estudio confirmatorio, por lo que puede considerarse que la información proporcionada por los índices CMIN y CMIN/df puede rechazarse por poco fiable.

Todo lo anteriormente descrito proporciona evidencias plausibles congruentes con la hipótesis de que el test evalúa un único factor relacionado con la habilidad en la conducción aquí definida como constructo.

Sensibilidad de la prueba

Ha sido señalado anteriormente que no se realizó un estudio pormenorizado de la validez predictiva del test, a pesar de que se disponía de datos referentes a la siniestralidad de los conductores de la Organización. Esto se debe al hecho de que estos datos carecían de una codificación uniforme y claramente definida, por lo que cabían dudas acerca de cómo se recababan, se codificaban o de cómo se señalizaba la responsabilidad del conductor y el tipo de siniestro en el que se había visto implicado.

A pesar de esto, en un estudio piloto, se trató de cruzar datos obtenidos de la prueba de evaluación de conducción con los datos recogidos en el historial de siniestros de un grupo de conductores. Sin embargo, la dificultad de codificación y la imposiblidad de establecer controles experimentales a variables tales como hora del día del siniestro, tipo de vehículo, tipo de ruta, jornada del conductor previa a la ocurrencia del accidente, experiencia previa en ese servicio, etcétera, recomendaron la no inclusión de resultado alguno referido al valor predictivo de la prueba de evaluación sobre la variable siniestralidad, entendida esta como riesgo futuro de sufrir un siniestro conduciendo un autobús.

Lo que si se pudo realizar a partir de esos datos fue un estudio de la sensibilidad de la prueba, esto es, de su capacidad para clasificar a los candidatos como buenos o malos conductores en cuanto a siniestros. La tabla de resultados del estudio de Regresión Logística realizado permite concluir que la prueba muestra una buena sensibilidad a la hora de clasificar a los candidatos de baja siniestralidad, con un 97% de clasificaciones correctas. Sin embargo, la sensibilidad de la prueba con sujetos de siniestralidad más elevada no es tan buena. Tan solo un 7% de estos candidatos son bien clasificados por el test.

Teniendo en cuenta la limitaciones, anteriormente expuestas, en los controles experimentales y la precariedad de los datos de siniestralidad obtenidos para este estudio, estos resultados deben observarse con cautela y tomarse como un estudio previo que marca una línea de trabajo en el desarrollo de investigaciones futuras, más profundas, sobre la capacidad predictiva del test como instrumento de selección de personal.

6.1.2 Fiabilidad

Desde la perspectiva clásica de los test se estimó la fiabilidad de la medida de la prueba de evaluación de conductores a partir de la medida de la consistencia interna del test. El coeficiente α refleja el grado en que correlacionan entre sí los ítems de un test, por lo tanto es esperable que en un test con una estructura unidimensional el nivel de α sea elevado, aunque el caso contrario no necesariamente se cumple siempre. Ya se comentó que un α elevado no presupone la unidimensionalidad del test.

Para el caso del presente test los dos cálculos de α que se realizaron dan resultados muy similares, ambos muy próximos a 0,90, en concreto, 0,886 utilizando el método RELIABILTY en SPSS y 0,899 usando la solución para el cálculo de α implementada en el programa FACTOR que se utilizó para el análisis factorial exploratorio.

Ambos valores son suficientemente elevados como para poder concluir que este test de evaluación de conductores de autobús es una herramienta fiable en su proceso de medida.

El estudio de TRI realizado tras el estudio clásico muestra también un buen dato de fiabilidad. En este caso el valor de fiabilidad marginal estimado es de 0,90. La función de información obtenida para el test revela que su mayor precisión de medida se da en niveles medios de la variable habilidad.

Estos datos reafirman la idea de que este es un test con una buena fiabilidad de medida.

El único punto que queda pendiente en cuanto al estudio de la fiabilidad del test, al tratarse de una prueba de metodología observacional, es la realización de un estudio de su fiabilidad interjueces, esto es, su capacidad de resistir diferentes evaluaciones realizadas por diferentes evaluadores.

Actualmente existe un manual de aplicación que es común a todo el personal que utiliza el test como herramienta en diferentes ámbitos, bien sea en pruebas de selección, como apoyo en procesos de formación o para evaluación de conductores en activo. A pesar de que en esta investigación no se realizó es factible pensar que, en estas condiciones, puede diseñarse un estudio adecuado que permita la resistencia del test a diferentes jueces evaluadores.

En el presente estudio los datos provenían de 35 evaluadores diferentes, principalmente fueron 6 de ellos los que aportaron el mayor número de pruebas a la muestra total, en concreto un 80%. El índice de consistencia interna tan elevado que se obtuvo, en esas condiciones, hace presuponer que la fiabilidad intejueces será también alta, ya que para este estudio se utilizaron diferentes evaluadores.

Sin embargo, lo apropiado será poner a prueba esa hipótesis en un estudio cuyo diseño contemple la evaluación de un mismo sujeto por diferentes evaluadores en condiciones similares y con los controles experimentales correspondientes y necesarios para poder obtener conclusiones relevantes y científicamente sólidas.

6.1.3 Análisis de ítems

Por la particularidad de que la escala en la que se responde al test es una escala de tipo Likert, la interpretación de los índices y parámetros estimados en el análisis de ítems difiere de la que se haría si los datos se hubieran obtenido a partir de un test al que se respondiera en otro tipo de escala, incluso si este fuera de acierto o fallo.

El *Índice de dificultad*, desde la perspectiva clásica de los test, cuando se calcula en una escala Likert no se considera como tal, sino que más bien expresa la probabilidad de dar una respuesta u otra en el continuo de la escala en función del nivel aptitudinal en la variable medida (Muñiz, et al., 2005).

En el caso del test estudiado en este trabajo los trece ítems del test tienen un índice de dificultad medio, mostrando bastante homogeneidad en las puntuaciones y siendo la mayoría de ellas próximas a la puntuación media de la escala de respuesta. El ítem que puede ser considerado como el más difícil es el ítem 2 *Uso del retarder*. Esto puede explicarse si se tiene en cuenta que el retardador hidráulico de frenada es uno de los instrumentos del vehículo autocar menos conocidos, de hecho su uso no se enseña, ni se exige, en la formación práctica obligatoria para la obtención de un permiso de clase D y, aunque existe como elemento de seguridad activa en los camiones, su uso y prestaciones son ligeramente diferentes a las del sistema montado en el autobús. Teniendo en cuenta que un alto porcentaje de la muestra está formada por conductores con menos de dos años de permiso (noveles o con escasa experiencia previa) y que, de aquellos que tienen experiencia previa, muchos la tienen con camión en el sector de transporte de mercancías, donde el uso de este instrumento no está tan

extendido, es fácilmente comprensible que las puntuaciones más bajas se obtengan en el ítem que, precisamente, mide esta habilidad.

A su vez, el ítem que se presenta como el más fácil, el ítem 9 *Atención a la señalización vial*, responde a una explicación igualmente lógica. No se debe olvidar que la normativa vial es perfectamente conocida por los conductores, máxime en los que se dedican profesionalmente al ejercicio del transporte, tanto por aquellos que acaban de obtener el permiso de conducción, como por aquellos experimentados que están poniéndola en práctica todos los días. Además la normativa es común para todos los conductores y no se dan especiales diferencias entre el autocar y el camión, salvo en las velocidades máximas y en algunas limitaciones de paso. Un último factor que influye en las altas puntuaciones obtenidas en este ítem es el hecho de que, al tratarse de una prueba de evaluación, inconscientemente el cumplimiento de la normativa por parte del candidato se hace de una manera especialmente rigurosa, sobre todo en el cumplimiento de las limitaciones de velocidad.

Desde la perspectiva de la Teoría de Respuesta al Ítem el parámetro *b* de dificultad también difiere al tratarse de una escala de respuesta graduada. En este caso, como ya se comentó anteriormente, el parámetro *b* expresa el nivel de habilidad θ a partir del cual es más probable que se cambie de categoría dentro de la escala de respuesta. Se considera entonces que más que un parámetro de dificultad se trata de un parámetro de localización.

En el estudio que se realizó sobre los ítems del test se ha comprobado que todos los parámetros de localización son bastante homogéneos situándose en el centro de la escala de habilidad. En este sentido, son los ítems 7 *Distancia de seguridad* y 9 *Atención a la señalización vial* los que requieren un nivel más bajo de θ para obtener puntuaciones más altas en el test. De la misma manera, es el ítem 11 *Observación y utilización del cuadro de mandos* el que requiere mayor nivel de habilidad para obtener puntuaciones altas, por lo que puede ser considerado el más difícil desde el análisis de TRI.

Por lo que se refiere a la capacidad de Discriminación de los ítems todos, desde la perspectiva clásica, presentan valores aceptables. El ítem 7 y el ítem 9 son los únicos que no alcanzan un valor de discriminación por encima de 0,50. Este resultado es congruente con el obtenido en el estudio de TRI, donde el parámetro *a* de casi todos los ítems presenta, según la escala de Baker (1985), una capacidad de discriminación alta o muy alta, esto es superior a 1,34. Únicamente los ítems 7 y 9 tienen valores

inferiores, considerándose su capacidad de discriminación como moderada en la escala de Baker.

En el caso de la distancia de seguridad (ítem 7), parece que el entorno desconocido, del desconocimiento del vehículo y de la ruta, así como la situación de evaluación, influyen en generar una actitud inicialmente más prudente, la cual lleva a mantener una mayor distancia de seguridad a todos los candidatos, lo que, independientemente de su experiencia previa o estilo de conducción, reduce la capacidad de discriminación del ítem.

Cuando se profundiza en la atención a la señalización vial (ítem 9), ya se ha señalado que la situación de evaluación que se produce al llevar detrás al evaluador experto, provoca una actitud más vigilante con la señalización vial que se traduce en un mayor cumplimiento de los límites de velocidad, mayor celo en las situaciones de paso y cruce de vehículos, etcétera. Nuevamente, este hecho se da de manera generalizada en todos los conductores independientemente de su experiencia o estilo de conducción.

6.2 Discusión de estudios específicos

Como complemento a la investigación principal sobre las propiedades psicométricas de la prueba se realizaron, tal y como se describió en los capítulos precedentes, una serie de estudios parciales para comprobar las hipótesis planteadas acerca de las relaciones o posibles influencias de determinadas variables sobre los resultados definitivos obtenidos en la prueba de evaluación.

Esta verificación de las hipótesis sobre la existencia de relaciones estadísticamente significativas o, en su caso, la comprobación de la independencia de los resultados de la prueba de otras variables ajenas a la habilidad en la conducción, mejora el futuro control experimental y favorece el conocimiento más profundo de las propiedades del test como herramienta de medida y sus posibles aplicaciones posteriores.

6.2.1 Edad

Los datos obtenidos a partir del Análisis de varianza realizado, teniendo en cuenta la variable Edad del candidato/a, confirman la existencia de una influencia de esta variable sobre los resultados de la prueba.

El estudio más pormenorizado de esta diferencia revela que es el grupo más joven el que muestra las diferencias estadísticamente significativas con respecto a las medias de otros grupos. En el grupo de edad inferior a 25 años las diferencias no solo son estadísticamente significativas sino que, además, muestra un mayor tamaño del efecto, siendo este moderado en la comparación con todos los grupos exceptuando el de 25-30.

Estos resultados parecen mostrar que el efecto de la Edad podría estar mostrando una influencia de la experiencia, enmascarada tras un supuesto *efecto edad*. Nótese que únicamente el grupo de menores de 25 presenta diferencias significativas con respecto al resto de los grupos y, como se ha visto, con un tamaño del efecto sensiblemente más grande. Aún pudiendo argumentarse que la Edad, en el ámbito de competencia de conductores profesionales, no implica necesariamente mayor experiencia al volante de un autobús (lo cual es categóricamente cierto, teniendo en cuenta que puede accederse a un permiso de la clase D en cualquier momento con el único requisito de estar, previamente, en posesión de un permiso clase B), debe tenerse en cuenta que si existe una mayor experiencia previa en la conducción de otro tipo de vehículos que mediatiza los resultados.

Toda vez que a partir de los 25 años en adelante (lo que supondría una experiencia de conducción máxima de siete años con vehículos de cuatro ruedas) no se observan diferencias relevantes debidas a la edad y que el núcleo fuerte de las diferencias se sitúa en el grupo más joven, es posible deducir que es la mayor experiencia y no la edad el factor que media en la obtención de mejores resultados en la prueba de evaluación.

6.2.2 Experiencia

A pesar de la intercorrelación con la que esta variable se presenta con la Edad (Maycock, et al., 1996) los resultados obtenidos en el estudio de Edad son congruentes con la idea de que el factor Edad por si mismo no es predictor de un mejor rendimiento en la prueba.

Sin embargo, cuando se analizan con detenimiento los resultados obtenidos en el estudio sobre la relación de la experiencia previa con vehículo pesado y los resultados de la prueba se observa que las diferencias son estadísticamente significativas para los diferentes grupos de manera más generalizada.

El grupo menos experimentado (<1 año) es el que muestra diferencias con todos los demás grupos, con un tamaño del efecto moderado en todos los casos. En los grupos de experiencia intermedia las diferencias se dan con los grupos con los que guarda más distancia, esto es, a mayor diferencia de años de experiencia las diferencias son significativamente mayores, también con un tamaño del efecto moderado.

El efecto positivo que la experiencia tiene sobre variables como la percepción del riesgo potencial (A. Borowsky, et al., 2010; Sagberg y Bjørnskau, 2006), el procesamiento de la información visual (Chapman y Underwood, 1998; Neyens y Boyle, 2008) o la alerta situacional (Kass, Cole, y Stanny, 2007) parecen influir en variables que forman parte de la evaluación de la prueba, como pueden ser la anticipación o el mantenimiento de la distancia de seguridad, lo que directamente redunda en puntuaciones mejores en el resultado total del test.

Además, cuando se controla el efecto de la variable Experiencia puede observarse que la variable Edad no es un predictor tan fiable y las diferencias significativas desaparecen, dándose la circunstancia de que son los candidatos más jóvenes y los de mayor edad los que obtienen peores puntuaciones.

Esto revela de manera evidente que es la Experiencia y no tanto la Edad la variable que afecta más positivamente a los resultados del test.

Se confirma así la hipótesis de que, si el test mide habilidad en la conducción, aquellos sujetos con mayor experiencia al volante deberían ser los que obtuvieran resultados significativamente mejores.

6.2.3 Sexo

Las diferencias en la conducción (técnica, seguridad, siniestralidad, etcétera) entre hombres y mujeres son, probablemente, uno de los temas más clásicos y recurrentes en la literatura especializada desde principios del siglo XIX.

Aún así lo relevante para esta investigación es comprobar si la prueba trata por igual a los candidatos independientemente de su género, esto es, verificar si la prueba presenta algún tipo de sesgo o funcionamiento diferencial motivado por el sexo del candidato/a y no por las diferencias propias de sus habilidades en la conducción.

El estudio de medias realizado revela que en esta muestra y para esta prueba existen diferencias significativas en las puntuaciones de hombres y mujeres, en perjuicio de estas últimas.

El estudio de DIF que se realizó a continuación de este descubrimiento explica la existencia de estas diferencias por la significación estadística, en el estudio de regresión logística, de ocho de los ítems que conforman el test.

Sin embargo, la combinación en el estudio del cálculo del tamaño del efecto de esas diferencias reveló que ninguno de los ítems del test mostraba funcionamiento diferencial entre hombres y mujeres, por lo que parece plausible atribuir esas diferencias significativas al elevado tamaño muestral que afecta al estadístico χ^2.

6.2.4 Atención

Se destacó anteriormente que los problemas atencionales tienen una relación directa con una alta siniestralidad en la conducción. Tal y como se explicó, esta vinculación viene dada por la presencia clave de los procesos atencionales en la correcta percepción de los estímulos.

Aún reconociendo esa presencia y la relevancia de la Atención en la seguridad vial, en el tema que se trata en este momento su relevancia no debería ser tal, o al menos, no tan clave. La pretensión ideal sería que la prueba de evaluación, al estar midiendo únicamente habilidad en la conducción, no esté influida por la mayor o menor capacidad atencional del candidato que se somete a la prueba.

Para verificar este punto se llevó a cabo un estudio específico tomando datos de la capacidad atencional de una muestra de candidatos que se sometían también a una evaluación mediante la prueba de evaluación de conductores. Esta medida de la atención se realizó utilizando un test clásico el d2 (Brickenkamp, 2009). La muestra de 106 sujetos fue obtenida dentro de la muestra general de candidatos que se presentaron a pruebas de selección para conductores.

Los resultados del análisis de varianza que se realizó posteriormente reflejan que no se aprecian diferencias significativas en los resultados de la prueba en función del nivel atencional del sujeto, ni cuando se tiene en cuenta la escala total, ni considerando las escalas parciales de la batería d2.

Así mismo, el análisis de tablas de contingencia que se realizó tras el análisis de varianza tampoco permite considerar la existencia de dependencia entre las variables Atención y Puntuaciones en la evaluación de la prueba de conducción.

Estos resultados son congruentes con la hipótesis de que no debería existir relación entre el nivel atencional del candidato y sus resultados en la prueba. Debe tenerse en cuenta que el objetivo de la prueba es medir Habilidad en la conducción,

por lo que no es esperable que el nivel atencional mejore o empeore los resultados de la evaluación.

Si se pretende predecir el riesgo de sufrir un accidente, por la relación que se ha demostrado que existe entre esa variable y el nivel atencional, no sería necesario el uso de una evaluación de habilidad en la conducción. Una medida válida y fiable de la capacidad atencional del sujeto podría utilizarse como elemento de juicio para determinar la probabilidad de padecer un accidente. En todo caso, podría abrirse una vía de investigación interesante combinando ambas medidas y utilizándolas como predictores tanto tomadas por separado como la combinación de ambas.

6.2.5 Personalidad

Nuevamente, la relación de esta variable con el tópico riesgo de padecer un accidente de tráfico está sobradamente estudiada y aparece sistemáticamente a lo largo de la literatura científica.

Aún así, tal y como sucede con la variable Nivel Atencional, debe esperarse la independencia de la puntuación en la prueba de evaluación de conductores de los diferentes perfiles de personalidad que pueden darse.

Como ya se señaló en el punto anterior al aludir a su relación con el nivel atencional, si la prueba evalúa habilidad en la conducción los resultados obtenidos deberían, en este caso, estar libres de la influencia del perfil de personalidad del candidato que se somete a ella.

De nuevo, podrá ser objeto de estudio si la combinación de un determinado nivel de habilidad en la conducción con un determinado tipo de perfil de personalidad puede ser mejor o peor predictor de riesgo de sufrir un accidente, pero no es esperable que las puntuaciones de la prueba de evaluación varíen en función de los diferentes tipos de personalidad.

Para verificar esta independencia se utilizaron los datos obtenidos en una submuestra de 101 sujetos, que se habían sometido a la evaluación de conducción, en una prueba de personalidad. Para ello se les pasó una forma abreviada del test NEO PI-R.

El análisis de varianza realizado con todos los datos disponibles no muestra diferencias estadísticamente significativas en las puntuaciones de la prueba de conducción, motivadas por las puntuaciones en el test de personalidad.

Estos resultados son congruentes con la hipótesis de que la evaluación de conducción realizada con el test es independiente del perfil de personalidad del candidato que se presenta.

Conclusiones y Potencialidades Futuras de la Herramienta

7

En este capítulo final se presentan las conclusiones más significativas extraídas de los estudios realizados a la prueba.

Así mismo se pretende también señalar posibles escenarios de uso de la herramienta de evaluación, presentes y futuras. También se marcan las líneas esenciales de las investigaciones que serán relevantes a partir de este punto para mejorar la calidad de la información extraída de los resultados de la prueba.

7.1 Conclusiones

La evaluación mediante test y pruebas psicométricas aplicada a la selección de personal es algo que se viene realizando sistemáticamente desde el mismo inicio de la evaluación psicológica. Uno de los hitos relevantes en el desarrollo de la Psicometría como disciplina está, precisamente, relacionado con la aplicación de test colectivos diseñados para seleccionar y clasificar a grandes grupos de personas, el ejemplo clásico son los test α y β realizados durante la Gran Guerra para seleccionar reclutas en el ejercito americano (García Cueto, 1993; Muñiz, 1992).

La combinación de diferentes teorías y presupuestos psicológicos con las técnicas y herramientas que las matemáticas y la estadística ponen a disposición de los expertos ha permitido el desarrollo de instrumentos de medida de las más diversas variables psicológicas. Tratándose siempre de una medida indirecta, los principales esfuerzos han caído del lado de desarrollar herramientas fiables y válidas, que permitieran la elaboración de inferencias a partir de los datos con un cierto grado de confiabilidad.

A ese respecto, el caso de la prueba de evaluación de conductores de autobús que se presenta aquí no es diferente. Se ha pretendido que el desarrollo de una herramienta utilizada para la selección y evaluación de personal cumpliera con los estándares mínimos exigidos para poder certificar que la medida es tanto fiable como válida.

Tras el estudio realizado pueden extraerse las siguientes conclusiones, relacionadas con los objetivos e hipótesis planteados inicialmente:

(I) En lo referente a evidencias basadas en el contenido del test, puede concluirse que el proceso de construcción de la herramienta ha sido

adecuado, tanto en la definición del constructo a evaluar como en la formulación de los ítems que sirven para su medida.

(II) La prueba es unidimensional. Mide un único constructo, definido este como *Habilidad en la conducción*

(III) El test presenta una alta consistencia interna.

(IV) Los valores de dificultad de los trece ítems son homogéneos y la mayoría de ellos presentan buenos índices de discriminación. Asimismo, los resultados del análisis de ítems desde la TRI son congruentes con la visión clásica. Puede concluirse que las propiedades de medida del test son adecuadas. El test mide adecuadamente el constructo para el que fue ideado.

(V) La variable *Edad* no tiene efectos sobre los resultados de la prueba.

(VI) Las variable *Experiencia* está relacionada con los resultados finales en la prueba de la siguiente manera: una mayor experiencia previa en conducción de vehículo pesado favorece una mejor puntuación final.

(VII) No se ha constatado que el test tenga un funcionamiento diferencial que perjudique sistemáticamente al grupo de mujeres evaluadas.

(VIII) Los estudios específicos sobre la influencia de la variable *Atención* no han mostrado ningún efecto sobre los resultados en el test.

(IX) Los resultados del test son independientes de los perfiles de la variable *Personalidad* estudiados en la presente investigación.

7.2 Posibilidades de uso

Esta herramienta se creó en un primer momento como un instrumento de evaluación que diera soporte a un aspecto concreto de un proceso de selección global para conductores de autobús.

La vigencia de ese uso es innegable, poco a poco el uso de la herramienta se ha generalizado y extendido a todas las áreas de la Organización propietaria de ella, y ya se utiliza para selección de personal, no solo en los centros de selección principales, sino que también se utiliza por instructores debidamente formados en otros puntos de la geografía de la empresa como elemento decisorio en los procesos de selección.

La lógica de la concepción de la herramienta muestra que este es el uso principal que se le puede dar. Fue concebida para ello y es la utilidad en la que más ha sido probada y de la que se disponen de mayor número de registros. Los resultados están siendo satisfactorios y la confianza en los datos proporcionados por la evaluación con este test ha crecido hasta el punto de ser dato clave en la decisión de la inclusión o no de un candidato en fases más avanzadas del proceso.

Sin embargo, parece claro que su consistencia en la medida, su facilidad de uso y posibilidades de aplicación hacen de esta una herramienta con muchas posibilidades de uso, más allá de la selección de personal para la que fue concebida originalmente.

En la formación oficial que se establece dentro de los Certificados de Profesionalidad, dentro de la especialidad *Conductor de autobús*, el uso de la prueba de evaluación de conductores tiene cabida de diversas formas. Su utilización al inicio del curso proporciona una perfecta información de la situación pre del conductor, quien, tras tres meses de formación en habilidades de conducción, al

finalizar el curso puede realizar una prueba final que proporcione una información post perfectamente comparable con la inicial y que permitiría evaluar la progresión del alumno, así como valorar la calidad del proceso formativo, la transferencia de conocimiento al puesto de conducción e introducir mejoras en todo el planteamiento del desarrollo de la formación.

Es además un instrumento perfecto para poder dar información de retorno a los alumnos de cómo se va desarrollando su progresión en el aprendizaje y permitiría poder trabajar directa e individualizadamente con cada alumno aquellos aspectos en los que su nivel fuera inferior. Trabajar con una herramienta objetiva de evaluación permite avanzar en la formación de una forma personalizada, aportando valor a la formación en si y favoreciendo la mejora de los resultados.

Además, en una empresa privada que presta sus servicios en un entorno competitivo, esta es una herramienta que puede utilizarse perfectamente como instrumento de evaluación continua. De nuevo se mostrará útil para evaluar el pre y el post de una intervención formativa concreta, pero podrá utilizarse perfectamente como elemento de evaluación antes situaciones concretas, por ejemplo para evaluar el motivo que causa que un conductor concreto muestre un elevado número de siniestros en un plazo de tiempo. También puede destinarse su uso como herramienta de evaluación continua, incluyendo su resultado en una hipotética evaluación periódica que se realice a los conductores de un determinado servicio o una determinada empresa.

En general, cualquier uso relacionado con la selección, evaluación y formación de conductores sería perfectamente adecuado para esta herramienta. Su información es completa y se ha demostrado que está relacionada con la habilidad en la conducción del sujeto que la realiza, lo que permite en todo momento detectar carencias sobre las que se puede intervenir con formación específica.

7.3 Futuras vías y proyectos de investigación

Llegados a este punto final en la memoria de tesis únicamente queda por dejar apuntadas las líneas futuras de investigación que contribuirían a la mejora de la herramienta y al aumento de sus posibilidades.

Sin ninguna duda el trabajo más relevante que debe hacerse desde este punto es el desarrollo de una investigación pormenorizada de la Validez, desde la óptica de la relación test criterio. Como herramienta de selección la prueba contribuye a tomar decisiones sobre la incorporación, o no, de un candidato o sobre la continuidad, o no, en la empresa de un conductor. Aunque ha quedado demostrado que la fiabilidad en la medida es muy buena y que se pueden aportar suficientes evidencias basadas en su contenido y su estructura interna, lo que nos permite asegurar que la prueba mide habilidad en la conducción, para la toma final de decisiones es muy importante asegurar una adecuada relación test-criterio. Debe investigarse si la prueba es capaz de clasificar correctamente a los individuos que se someten a ella en términos de mayor o menor riesgo de siniestralidad. Eso pasa, inevitablemente por profundizar en su relación con otras variables, quizá no tan clásicas, pero más actuales en su valor como predictoras de siniestralidad vial.

Es necesario poder afirmar que el test proporciona prueba sólida sobre la siniestralidad futura de un candidato. De esta forma, su capacidad como herramienta de selección mejoraría sensiblemente.

Por último, debe trabajarse en el desarrollo de un nuevo manual de aplicación que, basado en la experiencia de uso de los últimos años, pueda pulir los aspectos que aún son difícilmente evaluables y aquellas zonas oscuras sobre las que

hay dudas a la hora de puntuar o no está claro el criterio que debe ser utilizado. En este punto, plantearse introducir más exhaustivamente técnicas y modelos de metodología observacional mejoraría sensiblemente el uso y aplicabilidad de la herramienta.

Referencias bibliográficas

AERA, APA, y NCME. (1999). *Standards for Educational and Psychological testing*. Washington D.C.: American Psychological Association.

af-Wåhlberg, A. (2004a). Characteristics of low speed accidents with buses in public transport: Part I. *Accident Analysis and Prevention, 36*, 63-71.

af-Wåhlberg, A. (2004b). Characteristics of low speed accidents with buses in public transport: Part II. *Accident Analysis and Prevention, 36*, 63-71.

af-Wåhlberg, A. (2005). Differential accident involvement of bus drivers. En L. Dorn (Ed.), *Driver Behaviour and Training* (Vol. II, pp. 383-391). Aldershot: Ashgate.

af-Wåhlberg, A. (2007). Making bad look good: The scientific claims of Interactive Driving Systems. Obtenido de http://www.psyk.uu.se/digitalAssets/19/19252_af_Wahlberg_Making_bad_look_good.pdf

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A., y Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Albertsson, P., y Falkmer, T. (2005). Is there a pattern in European bus and coach incidents? A literature analysis with special focus on injury causation and injury mechanisms. *Accident Analysis and Prevention*(37), 225-233.

Álvarez, M. (2000). *Análisis estadístico con SPSS*. Bilbao: Universidad de Deusto.

Amstrong, J. C., y Soelberg, P. (1968). On the Interpretation of Factor Analysis. *Psychological Bulletin*(70), 361-364.

Andrews, E. C., y Westerman, S. J. (2012). Age differences in simulated driving performance: Compensatory processes. *Accident Analysis and Prevention, 45*, 660-668.

Anónimo. (1908a, 26/05/1908). Catástrofe automovilística. *Diario ABC,* pp. 10-11.

Anónimo. (1908b, 26/05/1908). Sucesos. *Diario ABC,* p. 8.

APA. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington D.C.: American Psychological Association.

Arbuckle, J. (2006). *Amos 7.0 user's guide*. Spring House, PA: Amos Development Co.

Arias, B. (2008). *Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS.* Comunicación presentada en el VI Simposio Científico SAID. Metodología en la investigación sobre discapacidad: Introducción al uso de las ecuaciones estructurales, Salamanca.

Baker, F. B. (1985). *The basics of item response theory*. Portsmouth, NH: Heinneman.

Balboni, G. (2006). Structural Equation Modeling: "Rules of Thumb" with Participants with Disabilities. En T. E. Scruggs y M. A. Mastropieri (Eds.), *Applications of Research Methodology (Advances in Learning and Behavioral Disabilities, Volume 19)* (pp. 83-100): Emerald Group Publishing Limited.

Ballester, R., Gil, M. D., Gómez, S., y Gil, B. (2010). Propiedades psicométricas de un instrumento de evaluación de la adicción al cibersexo. *Psicothema, 22*(4), 1048-1053.

Batista, J. M., y Coenders, G. (2000). *Modelos de ecuaciones estructurales* (Vol. 6). Madrid: La Muralla.

Bayam, E., Liebowitz, J., y Agresti, W. (2005). Older drivers and accidents: A meta analysis and data mining application on traffic accident data. *Expert Systems with Applications, 29*, 598.

Bechger, M., Maris, G., Verstralen, H., y Béguin, A. A. (2003). Using Classical Test Theory in combination with Item Response Theory. *Applied Psychological Measurement, 27*(5), 319-334.

Bentler, P. M. (1978). The interdependece of theory, methodology and empirical data. Causal modelling as an approach to construct validation. En D. B. Kandel (Ed.), *Longitudinal research on drug use. Empirical findings and methodological issues*. New York: Wiley.

Bergdahl, J. (2005). Sex differences in attitudes toward driving: A survey *The Social Science Journal, 42*, 595-601.

Bergdahl, J., y Norris, M. R. (2002). Sex differences in single vehicle fatal crashes: a research note. *The Social Science Journal, 39*, 287-293.

Birnbaum, A. (1968). Some latent trait models and their use in inferring a examinee's ability. En F. M. Lord y M. Novick (Eds.), *Statistical theories in mental test scores*. Reading: Addison Wesley.

Bloom, H., y Lipsey, M. W. (2004). Some food for thought about effect size. Obtenido de http:// www.wtgrantfoundation.org/resources/design-and-analysis-of-group-randomized-trials

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Bock, R. D., y Lieberman, M. (1970). Fitting a response model for no dichotomously scored items. *Psychometrika, 35*, 179-197.

Bollen, K. A., y Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA.: Sage.

Borowsky, A., Oron-Gilad, T., y Parmet, Y. (2009). Age and skill differences in classifying hazardous traffic scenes *Transportation Research Part F, 12*, 277-287.

Borowsky, A., Shinar, D., y Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention, 42*, 1240-1249.

Borowsky, A., Shinar, D., y Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving *Accident Analysis and Prevention, 42*, 1240-1249.

Bortkiewicz, L. (1898). *Das Gesetz der keinen Zahlen*. Teubner, Leipzig: B.G. Teubner.

Brickenkamp, R. (2009). *Manual Test de Atención d2* (3 ed.). Madrid: TEA Ediciones.

Brown, I. D. (1982). Exposure and experience are a confounded nuisance in research on driver behaviour. *Accident Analysis and Prevention, 14*(345-352).

Buja, A., y Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509-540.

Campbell, D. T., y Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Cantin, V., Lavallière, M., Simoneau, M., y Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis and Prevention, 41*, 763-771.

Carbonell, E., Bañuls, R., Cortés, M. T., y Sáiz, E. J. (1995). El comportamiento humano en la conducción: Modelos explicativos. En L. Montoro, E. Carbonell, J. Sanmartín y F. Tortosa (Eds.), *Seguridad Vial: Del Factor Humano a las Nuevas Tecnologías* (pp. 163-201). Madrid: Síntesis.

Carr, D. B., Barco, P. P., Wallendorf, M. J., Snellgrove, C. A., y Ott, B. R. (2011). Predicting Road Test Performance in Drivers with Dementia. *Journal of the American Geriatrics Society, 59*(11), 2112-2117.

Cattel, R. B. (1966). The scree test for the number of factors. *Multivariate behavioural research, 1*(2), 245-276.

Chapman, P. R., y Underwood, G. (1998). Visual search of driving situations: danger and experience. *Perception, 27*(8), 951-964.

Chipman, M. L., MacGregor, C. G., Smiley, A. M., y Lee-Gosselin, M. (1993). The role of exposure in comparisons of crash risk among different drivers and driving environments. *Accident Analysis and Prevention, 25*, 207-211.

Clarke, D. D., Ward, P., Bartle, C., y Truman, W. (2006). Young drivers accidents in the UK: the influence of age, experience, and time of day. *Accident Analysis and Prevention, 38*, 871-878.

Clarke, D. D., Ward, P., Bartle, C., y Truman, W. (2010). Killer crashes: Fatal road traffic accidents in the UK. *Accident Analysis and Prevention, 42*, 764-770.

Clarke, D. D., Ward, P. J., y Jones, J. (1998). Overtaking road accidents: Differences in manoeuvre as a function of driver age. *Accident Analysis and Prevention*(30), 455-467.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, N.J.: Erlbaum.

Costa, P., y McRae, R. (2008). *Manual profesional del Inventario de Personalidad NEO PI-R en Inventario Reducido NEO-FFI*. Madrid: TEA Ediciones.

Costello, A. B., y Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), Available online: http://pareonline.net/getvn.asp?v=10&n=17.

Cresswell, W. L., y Froggatt, P. (1962). Accident Proneness, or Variable Accident Tendency? *Journal of the Statistical and Social Inquiry Society of Ireland, XX*(Part 5), 152-171.

Crocker, L., y Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika, 16*, 297-334.

Crossen, C. (1994). *Tainted truth. The manipulation of fact in America*. Nueva York, NY: Simon and Schuster.

De Raedt, R., y Ponjaert-Kristoffersen, I. (2000). The relationship between cognitive/neuropsychological factors and car driving performance in older adults. *Journal of the American Geriatric Society, 48*(12), 1664-1668.

del Río, M. C., Prada, R., Martín, F., y Alvarez, F. J. (1991). Fármacos, alcohol, drogas de abuso y conducción de vehículos en España. *Revista Mapfre de Seguridad, 41*, 23-29.

DGT. (2007). *Accidentes con implicación de autobuses y autocares* (No. NIPO: 128-07-004-6). Madrid: Observatorio Nacional de Seguridad Vial.

DGT. (2011). *Las principales cifras de la siniestralidad en España 2010*. Madrid: Observatorio Nacional de Seguridad Vial (DGT).

DGT. (2012). *Anuario estadístico de accidentes 2011*. Madrid: Dirección General de Tráfico.

DGTT. (2011a). *Observatorio del transporte de viajeros por carretera: Concesiones del Estado*. Madrid: Dirección General de Transporte Terrestre.

DGTT. (2011b). *Observatorio del transporte de viajeros por carretera: Oferta y demanda*. Madrid: Dirección General de Transporte Terrestre.

Di Milia, L., Smolensky, M. H., Costa, G., Howarth, H. D., Ohayon, M. M., y Philip, P. (2011). Demographic factors, fatigue, and driving accidents: An examination of the published literature. *Accident Analysis and Prevention, 43*, 516-532.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

Directiva 76/914/CEE del Consejo, de 16 de diciembre de 1976, relativa al nivel mínimo de formación de determinados conductores de vehículos de transporte por carretera (1976).

Directiva 2003/59/CE del Parlamento Europeo y del Consejo de 15 de julio de 2003 relativa a la cualificación inicial y la formación continua de los conductores de determinados vehículos destinados al transporte de mercancías o de viajeros por carretera (2003).

Dobbs, A. R., Heller, R. B., y MacDonald, W. (1998). A comparative approach to indentify unsafe older drivers. *Accident Analysis and Prevention, 30*(3), 363-370.

Donorfio, L. K. M., D'Ambrosio, L. A., Coughlin, J. F., y Mohyde, M. (2009). To drive or not to drive, that isn't the question—the meaning of self-regulation among older drivers. *Journal of Safety Research, 40*, 221-226.

Dorn, L., y af-Wåhlberg, A. (2008). Work-related road safety: An analysis based on U.K. bus driver performance. *Risk Analysis, 28*(1), 25-35.

Duke, J., Guest, M., y Boggess, M. (2010). Age-related safety in professional heavy vehicle drivers: A literature review. *Accident Analysis and Prevention, 42*, 364.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ.: Prentice-Hall.

Eckermann, E. (2001). *World History of the Automobile*. Troy MI: SAE International.

Ellis, B. B., y Mead, A. D. (2008). Item analysis: Theory and practice using classical and modern test theory. En S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology*. Malden, MA.: Blackwell Publishers.

Elosua, P. (2006). Funcionamiento diferencial del item en la evaluación internacional PISA. Detección y comprensión. Obtenido el 14-10-2012, de http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_4.htm

Elvik, R., y Vaa, T. (2006). *El manual de medidas de Seguridad Vial*. Madrid: FITSA.

ETSC. (2000). *Visions, Targets and Strategies*. Obtenido de http://www.etsc.eu/documents/visions%202000.pdf.

ETSC. (2009). *2010 on the horizon. 3rd Road Safety PIN Report*. Bruselas: European Transport Safety Council.

ETSC. (2010a). *Road Safety Target in Sight: Making up for the lost time. 4th Road Safety PIN Report*: European Transport Safety Council.

ETSC. (2010b). *Road Safety Target in Sight: Making up for the lost time. 4th Road Safety PIN Report*: European Transport Safety Council.

ETSC. (2011). *2010 Road Safety Target Outcome: 100.000 fewer deaths since 2001. 5th Road Safety PIN Report*: European Transport Safety Council.

Evans, L. (1985). Human behavior feedback and traffic safety. *Human Factors, 27*, 555-576.

Evans, L. (1991). *Traffic safety and the driver*. New York, NY: VanNostrand Reinhold.

Evans, L., y Schwing, R. C. (1985). *Human behavior and traffic safety*. Nueva York: Plenum Press.

Ewen, R. B. (1998). *Personality: A topical approach*. Mahweh, NJ: Erlbaum.

Farmer, E., y Chambers, E. G. (1926). *A psychological study of individual differences in accident rates* (No. 38). Lodon: Industrial Fatigue Research Board.

Ferrando, P. J. (1996). Relaciones entre el análisis factorial y la teoría de respuesta a los ítems. En J. Muñiz (Ed.), *Psicometría*. Madrid: Universitas.

Flanagan, J. C. (1937). A note on calculating the standard error of measurement and reliability coefficients with the test scoring machine. *Journal of Applied Psychology, 23*(4), 529.

Fuller, R. (1984). A conceptualization of driving behavior as threat avoidance. *Ergonomics, 27*, 1139-1155.

Furnham, A., y Saipe, J. (1993). Personality correlates of convicted drivers. *Personality and Individual Differences, 14*(2), 329-336.

García, A., Martín, A., y Páez, J. (2006). Analysis of the driver's workplace in coaches related to passive safety. *International Journal of Vehicle Desing, 42*, 134-153.

García Cueto, E. (1993). *Introducción a la Psicometría*. Madrid: Siglo XXI Editores.

García Cueto, E., Gallo, P., y Miranda, R. (1998). Bondad de ajuste en el Análisis Factorial Confirmatorio. *Psicothema, 10*(3), 717-723.

García-Cueto, E. (1993). *Introducción a la Psicometría*. Madrid: Siglo XXI Editores.

Germain, J., Pinillos, J. L., García, E., y Aberasturi, N. L. (1970). La validez de unas pruebas selectivas para conductores. *Revista de Psicología General y Aplicada, XXV*(107), 1067-1114.

Germain, J., Pinillos, J. L., y Pascual, M. (1959). Las pruebas de selección de conductores y su validez: Nota preliminar. *Revista de Psicología General y Aplicada, 14*(50), 421-429.

Gierl, M. J., y McEwen, N. (1995). *Differential Item Functioning on the Alberta Educational Social Studies 30 Diploma Exams.* Comunicación presentada en el Annual meeting of the Canadian Society for Studies in Education, Ottawa, Ontario, Canada.

Gómez-Benito, J., Hidalgo, M. D., y Zumbo, B. D. (2013). Effectiveness of Combining Statistical Tests and Effect Sizes When Using Logistic Discriminant Function Regression to Detect Differential Item Functioning for Polytomous Items. *Educational and Psychological Measurement, 73*(5), 875-897.

Graham, J. D. (1982). On Wilde's theory of risk homeostasis. *Risk Analysis, 2*(235-237).

Greenwood, M., y Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of attacks of decease or repeated accidents. *Journal of the Royal Statistical Society, 83*, 255-279.

Gregersen, N. P., y Bjurulf, P. (1996). Young novice drivers: towards a model of their accident involvement. *Accident Analysis and Prevention, 28*(2), 229-241.

Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement, 6*, 427-439.

Guion, R. M. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement, 1*, 1-10.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255-282.

Haight, F. A. (1980). What causes accidents? - a semantic analysis. En L. Evans (Ed.), *Accident Causation Report SP-461* (pp. 51-54). Warrendale, PA: Society of Automotive Engineers.

Haight, F. A. (1986). Risk, especially risk of traffic accident. *Accident Analysis and Prevention, 18*, 359-366.

Hambleton, R. K. (1990). Item response theory: introduction and bibliography. *Psicothema, 2*(1), 97-107.

Harman, H. H. (1980). *Análisis Factorial Moderno*. Madrid: Saltés.

Harrison, E. L. R., y Fillmore, M. T. (2005). Are bad drivers more impaired by alcohol? Sober driving precision predicts impairment from alcohol in a simulated driving task. *Accident Analysis and Prevention, 37*, 882-889.

Hauer, E., y Persaud, B. N. (1983). A common bias in Before-and-After accident comparisons and its estimation. *Transportation Research Record, 905*, 164-174.

Hedges, L. V., y Olkin, I. O. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Henson, R. K., y Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*(66), 393-416.

Hidalgo Montesinos, M. D., Gómez Benito, J., y Padilla García, J. L. (2005). Regresión Logística: Alternativas en la detección del Funcionamiento Diferencial del Ítem. *Psicothema, 17*(3), 509-515.

Hilakivi, I., Veilahti, J., Asplund, P., Sinivuo, J., Laitinen, L., y Koskenvuo, K. (1989). A sixteen-factor personality test for predicting automobile driving accidents of young drivers. *Accid Anal Prev, 21*(5), 413-418.

Hoggarth, P. A., Innes, C. R. H., Dalrymple-Alford, J. C., Severinsen, J. E., y Jones, R. D. (2010). Comparison of a linear and a non-linear model for using sensory–motor, cognitive, personality, and demographic data to predict driving ability in healthy older adults. *Accident Analysis and Prevention, 42*, 1759-1768.

Holland, P. W., y Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. En H. Wainer y H. I. Braun (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ.: LEA.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*(30), 179-185.

Hoyes, T. W., y Glendon, A. I. (1993). Risk homeostasis: Issues for future research. *Safety Science, 16*, 19-33.

Hunt, L. A., Murphy, C. F., Carr, D., Duchek, J. M., Buckles, V., y Morris, J. C. (1997). Reliability of the Washington University Road Test. A performance-based assessment for drivers with dementia of the Alzheimer type. *Archives of Neurology, 54*(6), 707-712.

IRTAD. (2010). *IRTAD 2010 Road Safety Annual Report*. Paris: Organisation for Economic Co-operation and Development.

IRTAD. (2013). *IRTAD 2013 Road Safety Annual Report*. Paris: Organisation for Economic Co-operation and Development.

Iversen, H., y Rundmo, T. (2002). Personality, risky driving, and accident involvement among Norwegian drivers. *Personality and Individual Differences, 33*, 1251-1263.

Jiménez, J. E., Hernández, S., García, E., Díaz, A., Rodríguez, C., y Martín, R. (2012). Test de atención D2: Datos normativos y desarrollo evolutivo de la atención en educación primaria. *European Journal of Education and Psychology, 5*(1), 93-106.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. En L. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford Press.

Jonah, B. A. (1997). Sensation seeking and risky driving: a review and synthesis of the literature. *Accident Analysis and Prevention, 29*(5), 651-665.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Justiss, M. D. (2005). *Development of a behind-the-wheel driving performance assessment for older adults.* no publicada, University of Florida, Florida.

Justiss, M. D. (2005). *Development of a behind-the-wheel driving performance assessment for older adults.* no publicada, University of Florida, Florida.

Kalyoncouglu, S. F., y Tigdemir, M. (2008). Modelling of personality, attitudes and risky driving. *Proceedings of the institution of Civil Engineers, Transport 161*(TRI), 37-43.

Kass, S. J., Cole, K. S., y Stanny, C. J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F 10*, 321-329.

Kelley, T. L. (1939). Selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17-24.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. Newbury Park: Sage.

Konstantopoulos, P., Chapman, P., y Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident Analysis and Prevention, 42*, 827-834.

Kuder, G. F., y Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika, 2*, 151-160.

Lancaster, R., y Ward, R. (2002). *The contribution of individual factors to driving behavior: Implications for managing work-related road safety*.

Lardelli Claret, P., Luna del Castillo, J. D., Jiménez Mejías, E., Pulido Manzanero, J., Barrio Anta, G., García Martín, M., et al. (2011). Comparison of two methods to assess the effect of age and sex on the risk of car crashes. *Accident Analysis and Prevention, 43*, 1555-1561.

Lawton, R., Parker, D., Stradling, S. G., y Manstead, A. (1997). Predicting road traffic accidents: The role of social deviance and violations. *British Journal of Psychology, 88*, 249-262.

Li, G., Braver, E., y Chen, L. H. (2003). Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. *Accident Analysis and Prevention, 35*, 227-235.

López Pina, J. A. (2005). Ítems politómicos vs dicotómicos: Un estudio metodológico. *Anales de Psicología, 21* (2), 339-344.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Lorenzo-Seva, U., y Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers, 38*(1), 88-91.

Mantel, N., y Haenszel, W. (1959). Statistical aspect of the analysis of data from restrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Martin, P. L., Audet, T., Corriveau, H., Hamel, M., D'Amours, M., y Smeesters, C. (2010). Comparison between younger and older drivers of the effect of obstacle direction on the minimum obstacle distance to brake and avoid a motor vehicle accident. *Accident Analysis and Prevention, 42*, 1144-1150.

Massie, D. L., Campbell, K. L., y Williams, A. F. (1994). Traffic accident involvement rates by driver age and gender. *Accident Analysis and Prevention, 27*, 73-87.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Maycock, G., Lester, J., y Lockwood, C. R. (1996). *The accident liability of car drivers: The reliability of self report data.* (No. 219). Crowthorne, England: Transport Research Laboratory.

Maycock, G., Lockwood, C. R., y Lester, J. F. (1991). *The accident liability of car drivers* (No. 315). Crowthorne, England: Transport and Road Research Laboratory.

Mayhew, D. R., y Simpson, H. M. (1990). *New to the Road Young Drivers and Novice Drivers: Similar Problems and Solutions?* Ottawa: Traffic Injury Research Foundation, Ont.

Mayhew, D. R., Simpson, H. M., y Pak, A. (2003). Changes in collision rates among novice drivers during first months of driving. *Accident Analysis and Prevention*(35), 683-691.

McCrae, R. R., y Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*(52), 81-90.

McCrae, R. R., y Costa, P. T., Jr. (1990). *Personality in adulthood*. New York: Guilford.

McCrae, R. R., y Costa, P. T., Jr. (1999). *Manual del NEO PI-R (Inventario de Personalidad NEO Revisado)*. Madrid: TEA Ediciones.

McDonald, R. P. (1981). The Dimensionality of Tests and Ítems. *British Journal of Mathematical and Statistical Psychology, 34*, 100-107.

McGwin, G., y Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis and Prevention*(31), 181-198.

McGwin Jr., G., y Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis and Prevention, 31*, 181-198.

McKenna, F. (1982). The human factor in driving accidents: An overview of approaches and problems. *Ergonomics*(25), 867-877.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7* (2), 105-107.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement, 19*, 91-100.

Mellenbergh, G. J. (1996). Modelos para ítems politómicos de respuesta discreta. En J. Muñíz (Ed.), *Psicometría*. Madrid: Universitas.

Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational Measurement* (3ª Edición ed., pp. 13-104). New York: Macmillan.

Mira, E. (1928). Laboratorio de Psicología. *Anales de la Sección de Orientación Profesional de la Escuela de Trabajo, 1*(1), 58-98.

Mislevy, R. J., y Bock, R. D. (1990). *BILOG 3 Item analysis and test scoring with binary logistic models*. Mooresville: Scientific Software.

Molenaar, I. W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden, 3*, 145-164.

Montoro, L., Alonso, F., Esteban, C., y Toledo, F. (2000). *Manual de Seguridad Vial: El factor humano*. Barcelona: Ariel.

Montoro, L., Carbonell, E., y Tortosa, F. (1991). Psicología y Seguridad Vial en España. 70 años de Historia. *Revista de la Historia de la Psicología, 12*(2), 73-86.

Montoro, L., Lijarcio, J. I., Roca, J., y Puchades, R. (2007). *Seguridad vial para nuevos conductores*. Madrid: Dirección General de Tráfico (DGT).

Morales Vallejo, P. (2010). El tamaño del efecto (effect size): Análisis complementarios al contraste de medias. Obtenido de http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oDelEfecto.pdf

Morales Vallejo, P. (2011). El Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios. Obtenido de http://www.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf

Muñiz, J. (1990). *Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Madrid: Pirámide.

Muñiz, J. (1992). *Teoría Clásica de los Test*. Madrid: Pirámide.

Muñiz, J. (1996). Fiabilidad. En J. Muñíz (Ed.), *Psicometría*. Madrid: Universitas.

Muñiz, J. (1998). La medición de lo psicológico. *Psicothema, 10*(1), 1-21.

Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R., y Moreno, R. (2005). *Análisis de los ítems* (Vol. 30). Madrid: La Muralla.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*(14), 59-71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Murray, W., y Cuerden, A. (2006). *Virtual Risk Manager: Validation studies*: Interactive Driving Systems Research report.

Muthén, B., y Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*(171-189).

Muthén, B., y Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45* (19-30).

Näätänen, R., y Summala, H. (1974). A model for the role of motivational factors in drivers decision-making. *Accident Analysis and Prevention, 6*, 243-261.

Neyens, D. M., y Boyle, L. N. (2008). The influence of driver distraction on the severity of injuries sustained by teenage drivers and their passengers. *Accident Analysis and Prevention, 40*(1), 254-259.

Nogueira, R., Godoy, A., Gavino, A., y Romero, P. (2009). Propiedades psicométricas del Vancouver Obsessional Compulsive Inventory (VOCI) en dos muestras no clínicas de población española. *Psicothema, 21*(4), 646-651.

Norman, L. G. (1963). *Los accidentes de tráfico: Epidemiología y Prevención*. Ginebra: OMS.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Nyberg, A., y Gregersen, N. P. (2007). Practicing for and performance on drivers license tests in relation to gender differences in crash involvement among novice drivers. *Journal of Safety Research, 38*, 71-80.

OCDE. (1990). *Behavioural adaptations to changes in the road transport system*. París: Organización para la Cooperación y el Desarrollo Económico.

Odenheimer, G. L., Beaudet, M., Jette, A. M., Albert, M. S., Grande, L., y Minaker, K. L. (1994). Performance-based driving evaluation of the eldery driver: Safety, reliability, and validity. *Journal of Gerontology, 49*(4), 153-159.

Oltedal, S., y Rundmo, T. (2006). The effect of personality and gender on risky driving behavior and accident involvement. *Safety Science, 44*, 621-628.

OSTC. (2010). *Observatorio Social del Transporte por Carretera 2009* (No. Noviembre 2010). Madrid: Ministerio de Fomento, Secretaría General de Transportes.

OSTC. (2011). *Oferta y Demanda* (No. Marzo 2011). Madrid: Ministerio de Fomento, Secretaría General de Transportes.

Paz, M. D. (1996). Validez. En J. Muñíz (Ed.), *Psicometría*. Madrid: Univesitas.

Pérez-Gil, J. A., Chacón Moscoso, S., y Moreno Rodríguez, R. (2000). Validez de Constructo: El uso del Análisis Factorial Exploratorio-Confirmatorio para obtener evidencias de validez. *Psicothema, 12*(2), 442-446.

R_Core_Team. (2013). R: A language and environment for statistical computing. Obtenido, de http://www.R-project.org/

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhague: The Danish Institute for Educational Research.

Real Decreto 170/2010, de 19 de febrero, por el que se aprueba el Reglamento de centros de reconocimiento destinados a verificar las aptitudes psicofísicas de los conductores. (2010).

Real Decreto 818/2009, de 8 de mayo, por el que se aprueba el Reglamento General de Conductores (2009).

Real Decreto 1032/2007 de 20 de julio, por el que se regula la cualificación inicial y la formación continua de los conductores de determinados vehículos destinados al transporte por carretera (2007).

Real Decreto 1467/1982, de 28 de mayo, por el que se determinan las enfermedades y deficiencias que pueden impedir la obtención o revisión de los permisos de conducción o restricciones a los mismos y las condiciones que deben reunir los certificados y reconocimientos correspondientes. (1982).

Real Decreto 1561/1995, de 21 de septiembre, sobre jornadas especiales de trabajo. (1995).

Real Decreto 1998/1996, de 6 de Septiembre, por el que se establece el Certificado de Profesionalidad de la Ocupacion de Conductor de autobus (1996).

Real Decreto 2272/1985, de 4 de diciembre, por el que se determinan las aptitudes psicofísicas que deben poseer los conductores de vehículos y por el que se regulan los centros de reconocimiento destinados a verificarlas. (1986).

Real Decreto Legislativo 339/1990, de 2 de marzo, por el que se aprueba el Texto Articulado de la Ley sobre Tráfico, Circulación de Vehículos a Motor y Seguridad Vial. (1990).

Recio, P., Cuadrado, I., y Ramos, E. (2007). Propiedades psicométricas de la escala de detección de sexismo en adolescentes (DSA). *Psicothema, 19*(3), 522-528.

Reglamento (CE) nº 561/2006, del Parlamento Europeo y del Consejo, de 15 de marzo de 2006, relativo a la armonización de determinadas disposiciones en materia social en el sector de los transportes por carretera (2006).

Reglamento (CE) No 2135/98 del Consejo de 24 de septiembre de 1998 por el que se modifica el Reglamento (CEE) no 3821/85 relativo al aparato de control en el sector de los transportes por carretera y la Directiva 88/599/CEE relativa a la aplicación de los Reglamentos (CEE) nos 3820/85 y 3821/85 (1998).

Reglamento (CEE) No 3821/85 del Consejo de 20 de diciembre de 1985 relativo al aparato de control en el sector de los transportes por carretera (1985).

Richardson, E. D., y Marottoli, R. A. (2003). Visual attention and driving behaviors among community-living older persons. *Journal of Gerontology A, 58*(9), 832-836.

Rockwell, T. (1981). Skills, judgement and information acquisition in driving. En T. Forbes (Ed.), *Human Factors in Highway Traffic Safety Research*. Florida: Robert E. Krieger.

Rodríguez, J. L. (1987). Los albores de la automoción en España: Un siglo sobre ruedas. *Revista Tráfico, 19,* 42-49.

Roe, C. M., Xiong, C., Shah, K., Morris, J. C., Duchek, J. M., Carr, D. B., et al. (2003). Longitudinal driving performance in early-stage dementia of the Alzheimer type. *Journal of the American Geriatrics Society, 51* (10), 1342-1347.

Romero, A., García Sevilla, J., Martínez Sánchez, F., Morales, A., y Sánchez Meca, J. (1990). Conduccion y vejez: el deterioro con la edad en tareas de velocidad de anticipacion y coordinacion visomotriz. *Anales de Psicología, 6*(2), 221-231.

Rosenbloom, T., Shahar, A., Perlman, A., Estreich, D., y Kirzner, E. (2007). Success on a practical driver's license test with and without the presence of another testee. *Accident Analysis and Prevention, 39*, 1296-1301.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review, 9*, 99-103.

Ryan, G. A., Legge, M., y Rosman, D. (1998). Age related changes in drivers' crash risk and chash type. *Accident Analysis and Prevention*(30), 379-387.

Sagberg, F., y Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers *Accident Analysis and Prevention, 38*, 407-414.

Samejima, F. (1969). Estimation of latent ability using a response pattern of grade scores. *Psychometric Monographs*(nº17).

Schwebel, D. C., Ball, K. K., Severson, J., Barton, B. K., Rizzo, M., y Viamonte, S. (2007). Individual difference factors in risky driving among older adults. *Journal of Safety Research, 38*, 501-509.

Scialfa, C. T., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., y Wetton, M. (2012). The effects of driving experience on responses to a static hazard perception test. *Accident Analysis and Prevention, 45*, 547-553.

Silva, F. (1989). *Evaluación conductual y criterios psicométricos*. Madrid: Pirámide.

Sinar, E. F., y Zickar, M. J. (2002). Evaluating the robustness of graded response model and classical test theory parameter estimates to deviant items. *Applied Psychological Measurement, 26*(2), 181-191.

Storie, V. J. (1977). *Male and female car drivers: Differences observed in accidents* (No. 761). Crowthorne, England: Transport and Road Research Laboratory.

Swaminathan, H., y Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370.

Tejedor, F. J. (1999). *Análisis de varianza* (Vol. 3). Madrid: La Muralla.

Thissen, D., Chen, W.-H., y Bock, R. D. (2003). *Multilog (version 7) [Computer software]*. Lincolnwood, IL: Scientific Software International.

Thomas, D. R., y Zumbo, B. D. (1996). *Variable importance in regression and related analysis.* Comunicación presentada en el Annual meeting of the Psychometric Society, Banff, AB, Canada.

Thorndike, R. L. (1951). Reliability. En E. L. Lindquist (Ed.), *Educational Measurement* (pp. 560-620). Washington, DC: American Council on Education.

Thurstone, L. L. (1935). *The vectors of the mind*. Chicago: University of Chicago Press.

Timmerman, M. E., y Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods*(16).

Torres, E., y Cuesta, M. (1992). 1992 Una revisión de las propiedades psicométricas del test de matrices progresivas de Raven (Escala superior). *Psicothema, 4*(1), 261-267.

Tortosa, F., Esteban, C., Tejero, P., y Molero, C. (1995). La selección psicológica de los conductores en España: Desarrollo histórico y situación del modelo actual. En L. Montoro, E. Carbonell, J. Sanmartín y F. Tortosa (Eds.), *Seguridad Vial: Del Factor Humano a las Nuevas Tecnologías* (pp. 375-395). Madrid: Síntesis.

Tortosa, F., y Montoro, L. (2002). La Psicología aplicada a la selección de conductores. Cien años salvando vidas. *Psicothema, 14*(4), 714-725.

Underwood, G., Chapman, P., Berger, Z., y Crundall, D. (2003). Driving experience, attentional focusing, and the recall of recently inspected events. *Transportation Research Part F: Traffic Psychology and Behaviour, 6*(4), 289-304.

Van der Linden, W. J., y Hambleton, R. K. (1997). *Handbook of modern item theory*. New York: Springer Verlag.

van Der Molen, H. H., y Bötticher, A. M. T. (1988). A hiearchical risk model for traffic participants. *Ergonomics, 31*(4), 537-555.

Van Elslande, P., y Faucher-Alberton, L. (1996). *When expectancies become certainties: a potential adverse effect of experience.* Comunicación presentada en el International Conference on Traffic and Transport Psychology, Valencia.

Vicente, M. T., Martín, A., Grimaldi, R., y Aparicio, F. (2008). *Nuevos requisitos y avances en seguridad pasiva en autobuses y autocares: Situación actual y necesidades.* Comunicación presentada en el 1er Congreso Iberoamericano de Seguridad Vial, CISEV, San José de Costa Rica.

Vrkljan, B. H., McGrath, C. E., y Letts, L. J. (2011). Assessment tools for evaluating fitness to drive: A critical appraisal of evidence. *Canadian Journal of Occupational Therapy, 78*(2), 80-96.

WIlde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis, 2*, 209-225.

WIlde, G. J. S. (1988). Risk homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissemsion in recent reactions. *Ergonomics, 31*, 441-468.

Yalow, E. S., y Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher, 12*(8), 10-14.

Yela, M. (1966). Jerarquías factoriales ortogonales y oblicuas. *Revista de Psicología General y Aplicada, 22*, 405-416.

Ziel, R. E. (2003). Maravillas modernas: Accidentes de tráfico: A&E Television Network.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, Ontario, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Apéndice: Anexos, Tablas y Análisis estadísticos

8.1 Salida completa de indicadores de ajuste del Análisis Factorial Confirmatorio

Analysis Summary

Notes for Group (Group number 1)

The model is recursive.
Sample size = 3857

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	45	785,957	59	,000	13,321
Saturated model	104	,000	0		
Independence model	26	18451,639	78	,000	236,559

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,957	,944	,960	,948	,960
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,756	,724	,726
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

NCP

Model	NCP	LO 90	HI 90
Default model	726,957	640,155	821,191
Saturated model	,000	,000	,000

Model	NCP	LO 90	HI 90
Independence model	18373,639	17930,396	18823,165

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	,204	,189	,166	,213
Saturated model	,000	,000	,000	,000
Independence model	4,785	4,765	4,650	4,882

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,057	,053	,060	,001
Independence model	,247	,244	,250	,000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	875,957	876,285		
Saturated model	208,000	208,758		
Independence model	18503,639	18503,828		

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	,227	,205	,252	,227
Saturated model	,054	,054	,054	,054
Independence model	4,799	4,684	4,915	4,799

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	383	428
Independence model	21	23

8.2 Resultados Multilog y Curvas Características de los items

RESULTADOS POR ÍTEM

```
MULTILOG for Windows 7.00.2327.2


 ITEM   1:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         1    2.04  (0.09)
 B( 1)     2   -0.70  (0.03)
 B( 2)     3    1.14  (0.05)
 B( 3)     4    6.82  (****)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.037  0.056  0.083  0.122  0.178  0.256  0.361  0.495
 -1.4 -  0.0  0.651  0.815  0.958  1.048  1.066  1.018  0.937  0.863
  0.2 -  1.6  0.830  0.851  0.918  1.002  1.060  1.058  0.983  0.850
  1.8 -  3.0  0.688  0.528  0.389  0.278  0.194  0.134  0.092

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3      4
  OBS. FREQ.    996   1541    458      0
  OBS. PROP.  0.3326 0.5145 0.1529 0.0000
  EXP. PROP.  0.2963 0.5121 0.1916 0.0000


 ITEM   2:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         5    1.85  (0.09)
 B( 1)     6   -0.53  (0.04)
 B( 2)     7    1.45  (0.06)
 B( 3)     8    7.34  (****)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.035  0.050  0.071  0.101  0.143  0.199  0.273  0.366
 -1.4 -  0.0  0.477  0.599  0.718  0.815  0.871  0.878  0.843  0.787
  0.2 -  1.6  0.734  0.706  0.714  0.754  0.812  0.862  0.881  0.853
  1.8 -  3.0  0.778  0.669  0.547  0.428  0.324  0.240  0.174

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3      4
  OBS. FREQ.    854   1378    391      0
  OBS. PROP.  0.3256 0.5254 0.1491 0.0000
  EXP. PROP.  0.3482 0.5057 0.1460 0.0000
```

```
ITEM   3:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A         9    2.35  (0.09)
B( 1)    10   -0.78  (0.03)
B( 2)    11    1.44  (0.04)
B( 3)    12    6.34  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.030  0.047  0.074  0.117  0.183  0.281  0.422  0.612
-1.4 -  0.0  0.846  1.093  1.295  1.387  1.338  1.174  0.967  0.785
 0.2 -  1.6  0.676  0.661  0.741  0.903  1.109  1.295  1.387  1.339
 1.8 -  3.0  1.166  0.926  0.684  0.478  0.322  0.212  0.137

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1114   2285    426      0
 OBS. PROP.  0.2912 0.5974 0.1114 0.0000
 EXP. PROP.  0.2654 0.6105 0.1241 0.0000


ITEM   4:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        13    1.65  (0.07)
B( 1)    14   -0.78  (0.04)
B( 2)    15    1.88  (0.08)
B( 3)    16    9.17  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.067  0.091  0.123  0.164  0.217  0.282  0.359  0.443
-1.4 -  0.0  0.528  0.604  0.659  0.684  0.677  0.641  0.587  0.528
 0.2 -  1.6  0.479  0.449  0.442  0.461  0.502  0.557  0.615  0.662
 1.8 -  3.0  0.685  0.676  0.635  0.568  0.486  0.401  0.320

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1140   2240    340      0
 OBS. PROP.  0.3065 0.6022 0.0914 0.0000
 EXP. PROP.  0.2943 0.6065 0.0992 0.0000


ITEM   5:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        17    2.09  (0.08)
B( 1)    18   -0.92  (0.03)
B( 2)    19    1.11  (0.04)
B( 3)    20    7.06  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.056  0.084  0.124  0.183  0.266  0.377  0.520  0.688
-1.4 -  0.0  0.861  1.008  1.093  1.095  1.023  0.914  0.810  0.751
 0.2 -  1.6  0.754  0.818  0.923  1.032  1.098  1.089  0.997  0.847
 1.8 -  3.0  0.673  0.507  0.367  0.258  0.178  0.121  0.082

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.    972   2105    689      0
 OBS. PROP.  0.2581 0.5589 0.1830 0.0000
 EXP. PROP.  0.2385 0.5641 0.1974 0.0000


ITEM   6:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        21    2.00  (0.08)
```

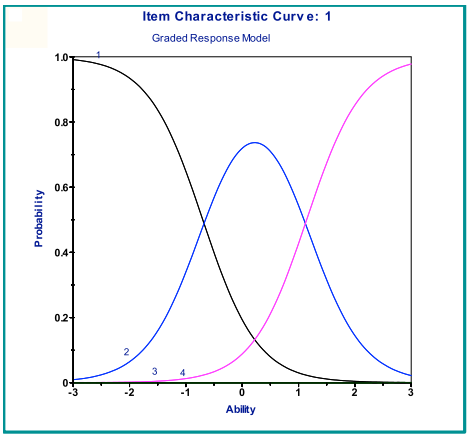
```
B( 1)    22   -0.74  (0.03)
B( 2)    23    1.50  (0.05)
B( 3)    24    7.41  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.043  0.063  0.092  0.135  0.195  0.277  0.384  0.518
-1.4 -  0.0  0.670  0.821  0.944  1.010  1.003  0.932  0.827  0.722
 0.2 -  1.6  0.650  0.628  0.661  0.741  0.849  0.950  1.010  1.003
 1.8 -  3.0  0.924  0.792  0.639  0.490  0.361  0.259  0.182

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1180   2187    446      0
 OBS. PROP.  0.3095 0.5736 0.1170 0.0000
 EXP. PROP.  0.2868 0.5842 0.1290 0.0000


ITEM   7:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        25    1.11  (0.06)
B( 1)    26   -1.94  (0.10)
B( 2)    27    2.26  (0.16)
B( 3)    28    8.54  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.221  0.247  0.269  0.288  0.301  0.308  0.308  0.301
-1.4 -  0.0  0.288  0.272  0.253  0.233  0.216  0.200  0.189  0.182
 0.2 -  1.6  0.181  0.185  0.193  0.206  0.223  0.241  0.261  0.279
 1.8 -  3.0  0.294  0.305  0.309  0.307  0.297  0.282  0.262

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.    586   2768    376      0
 OBS. PROP.  0.1571 0.7421 0.1008 0.0000
 EXP. PROP.  0.1457 0.7428 0.1113 0.0001


ITEM   8:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        29    1.53  (0.06)
B( 1)    30   -0.97  (0.04)
B( 2)    31    1.20  (0.06)
B( 3)    32    9.03  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.096  0.126  0.165  0.212  0.268  0.332  0.401  0.470
-1.4 -  0.0  0.530  0.576  0.602  0.606  0.592  0.569  0.545  0.530
 0.2 -  1.6  0.529  0.541  0.564  0.588  0.605  0.604  0.583  0.540
 1.8 -  3.0  0.481  0.414  0.345  0.279  0.222  0.173  0.133

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1000   1981    754      0
 OBS. PROP.  0.2677 0.5304 0.2019 0.0000
 EXP. PROP.  0.2588 0.5301 0.2111 0.0000


ITEM   9:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        33    0.95  (0.08)
B( 1)    34   -1.30  (0.07)
B( 2)    35    1.34  (0.12)
B( 3)    36   11.74  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
```

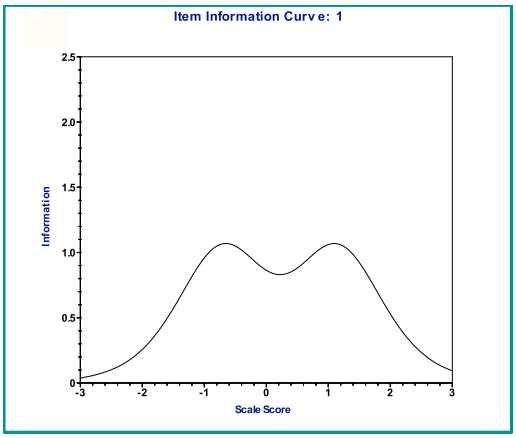
```
-3.0 - -1.6  0.125  0.142  0.158  0.175  0.191  0.206  0.218  0.229
-1.4 -  0.0  0.236  0.241  0.244  0.245  0.244  0.243  0.242  0.241
 0.2 -  1.6  0.242  0.243  0.244  0.245  0.244  0.242  0.237  0.230
 1.8 -  3.0  0.220  0.208  0.194  0.178  0.162  0.145  0.128

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1039   1841    900      0
 OBS. PROP.  0.2749 0.4870 0.2381 0.0000
 EXP. PROP.  0.2596 0.4865 0.2539 0.0000


ITEM  10:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        37    2.72  (0.10)
B( 1)    38   -0.83  (0.03)
B( 2)    39    0.96  (0.03)
B( 3)    40    6.01  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.020  0.035  0.059  0.101  0.170  0.284  0.462  0.725
-1.4 -  0.0  1.073  1.458  1.765  1.867  1.725  1.438  1.163  1.018
 0.2 -  1.6  1.053  1.256  1.552  1.803  1.858  1.668  1.316  0.935
 1.8 -  3.0  0.616  0.387  0.235  0.141  0.083  0.050  0.030

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1007   2054    741      0
 OBS. PROP.  0.2649 0.5402 0.1949 0.0000
 EXP. PROP.  0.2422 0.5472 0.2107 0.0000


ITEM  11:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        41    1.84  (0.08)
B( 1)    42   -0.17  (0.03)
B( 2)    43    1.45  (0.05)
B( 3)    44   13.90  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.019  0.027  0.038  0.055  0.078  0.110  0.154  0.214
-1.4 -  0.0  0.291  0.387  0.499  0.620  0.735  0.825  0.878  0.889
 0.2 -  1.6  0.870  0.842  0.826  0.833  0.858  0.884  0.887  0.852
 1.8 -  3.0  0.775  0.668  0.548  0.431  0.327  0.242  0.176

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1503   1379    487      0
 OBS. PROP.  0.4461 0.4093 0.1446 0.0000
 EXP. PROP.  0.4492 0.4053 0.1455 0.0000


ITEM  12:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        45    2.35  (0.09)
B( 1)    46   -0.46  (0.03)
B( 2)    47    1.00  (0.04)
B( 3)    48    6.63  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.014  0.022  0.035  0.056  0.089  0.140  0.217  0.331
-1.4 -  0.0  0.492  0.702  0.948  1.190  1.367  1.430  1.382  1.284
 0.2 -  1.6  1.217  1.229  1.312  1.404  1.425  1.327  1.125  0.876
 1.8 -  3.0  0.637  0.441  0.294  0.192  0.123  0.079  0.050
```

```
 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1235   1311    496      0
 OBS. PROP.  0.4060 0.4310 0.1631 0.0000
 EXP. PROP.  0.3559 0.4334 0.2106 0.0000


ITEM  13:       4 GRADED CATEGORIES
       P(#) ESTIMATE (S.E.)
A        49    3.05  (0.11)
B( 1)    50   -0.48  (0.03)
B( 2)    51    0.84  (0.03)
B( 3)    52    6.82  (****)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.004  0.008  0.014  0.026  0.048  0.088  0.158  0.283
-1.4 -  0.0  0.495  0.829  1.301  1.841  2.260  2.354  2.130  1.831
 0.2 -  1.6  1.719  1.881  2.189  2.365  2.199  1.738  1.200  0.753
 1.8 -  3.0  0.445  0.253  0.141  0.078  0.043  0.023  0.013

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2      3      4
 OBS. FREQ.   1175   1465    816      0
 OBS. PROP.  0.3400 0.4239 0.2361 0.0000
 EXP. PROP.  0.3401 0.4255 0.2344 0.0000


ITEM  14: GRP1, N[MU:  0.00 SIGMA:  1.00]
   P(#);(S.E.):   54; (0.00)   55; (0.00)

@THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
-1.4 -  0.0  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
 0.2 -  1.6  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
 1.8 -  3.0  1.000  1.000  1.000  1.000  1.000  1.000  1.000


TOTAL TEST INFORMATION

@THETA:      INFORMATION:
-3.0 - -1.6  1.766  1.995  2.307  2.736  3.327  4.140  5.239  6.674
-1.4 -  0.0  8.439 10.405 12.279 13.641 14.116 13.617 12.500 11.413
 0.2 -  1.6 10.932 11.288 12.279 13.337 13.841 13.515 12.482 11.018
 1.8 -  3.0  9.382  7.774  6.326  5.104  4.122  3.359  2.779

@THETA:      POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6  0.752  0.708  0.658  0.605  0.548  0.491  0.437  0.387
-1.4 -  0.0  0.344  0.310  0.285  0.271  0.266  0.271  0.283  0.296
 0.2 -  1.6  0.302  0.298  0.285  0.274  0.269  0.272  0.283  0.301
 1.8 -  3.0  0.326  0.359  0.398  0.443  0.493  0.546  0.600

MARGINAL RELIABILITY:    0.9050

NEGATIVE TWICE THE LOGLIKELIHOOD=      9095.4
(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)


 NORMAL PROGRAM TERMINATION

 START DATE: 09-25-2011
 START TIME: 00:10:19
 END   TIME: 00:10:21
```

CURVA CARACTERÍSTICA Y FUNCIÓN DE INFORMACIÓN DE LOS ITEMS

ÍTEM 1

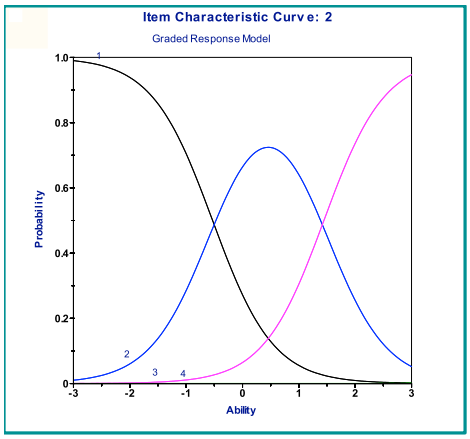


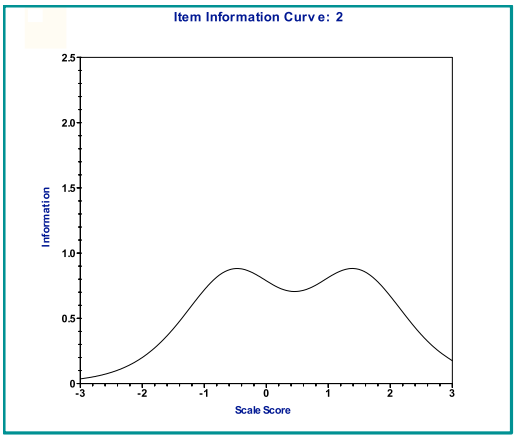


Category legends **Item: 1**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 2

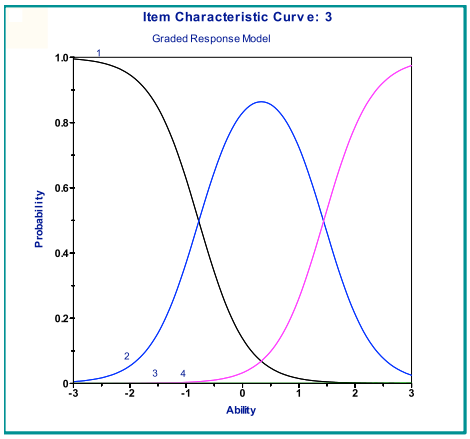


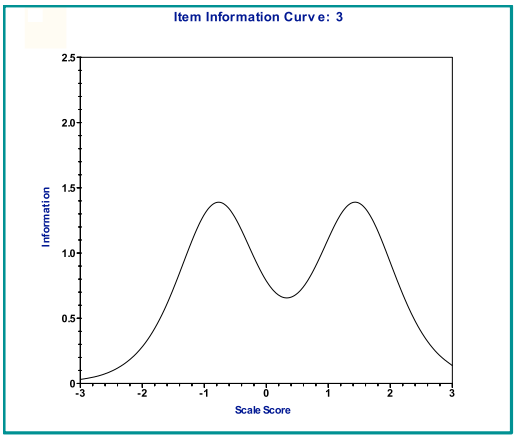


Category legends **Item: 2**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 3

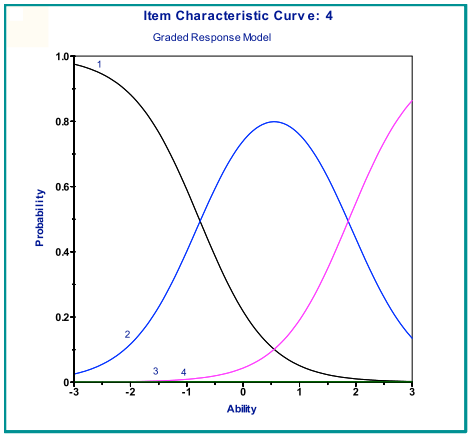


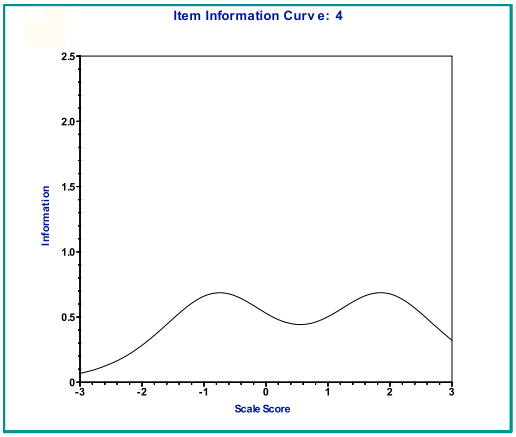


Category legends **Item: 3**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 4

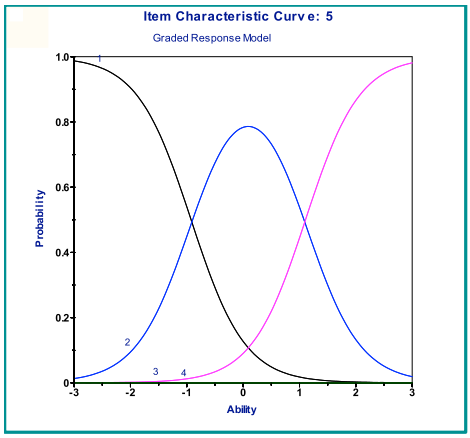


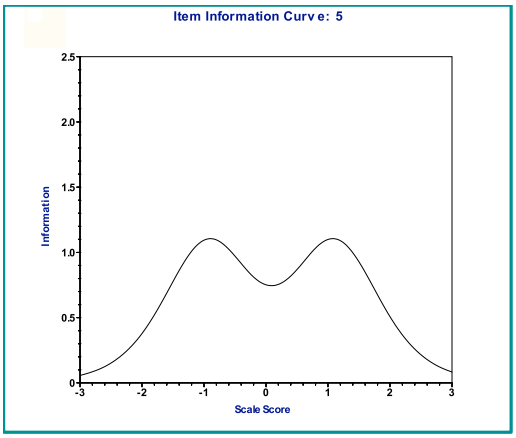


Category legends **Item: 4**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 5

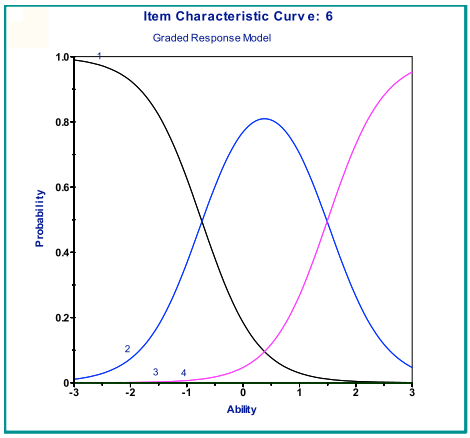


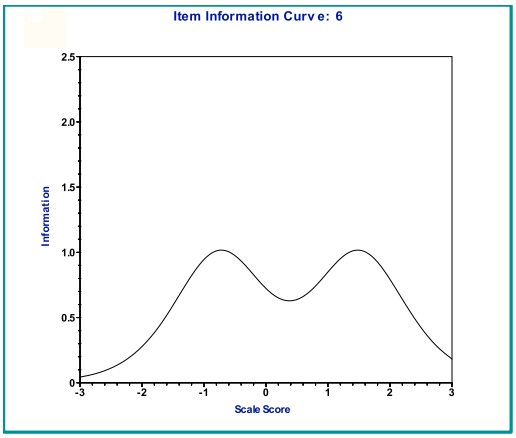


Category legends **Item: 5**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 6

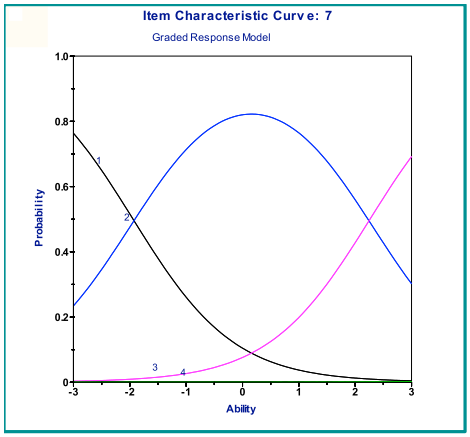


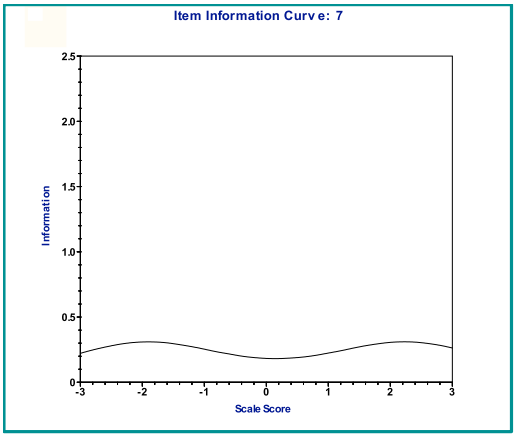


Category legends **Item: 6**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 7

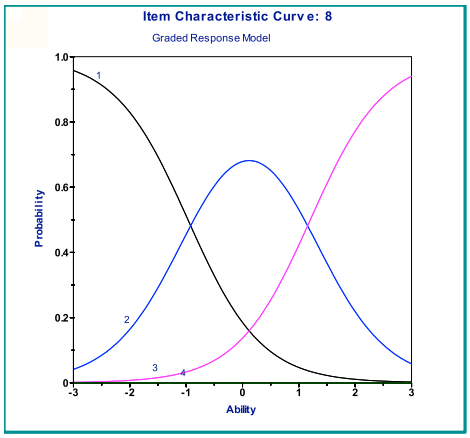


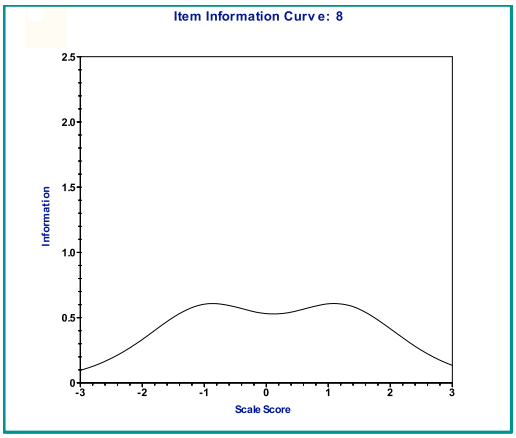


Category legends **Item: 7**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 8

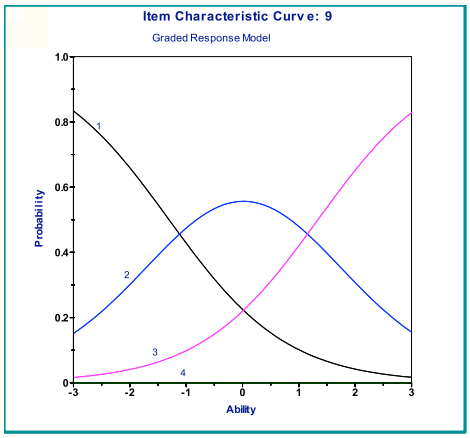


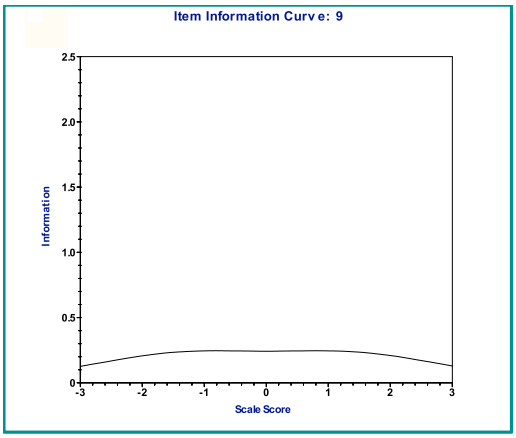


Category legends **Item: 8**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 9

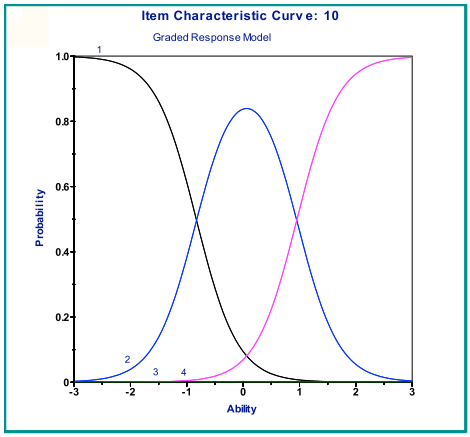


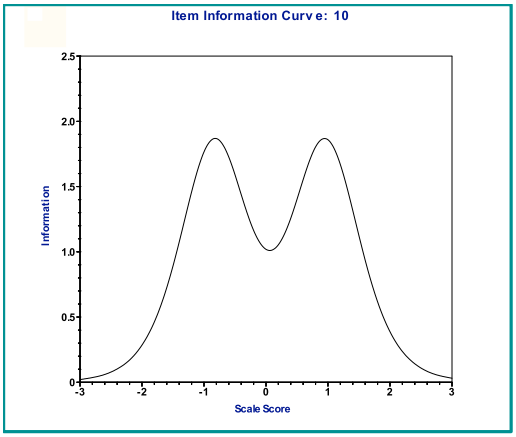


Category legends **Item: 9**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 10

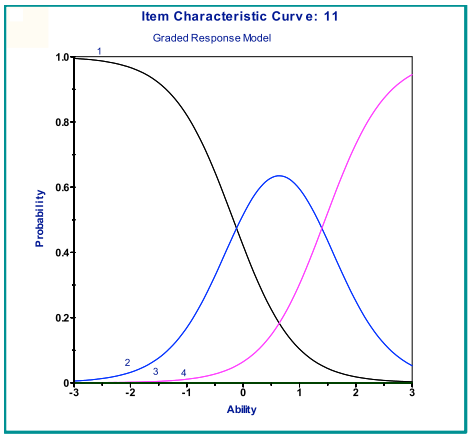


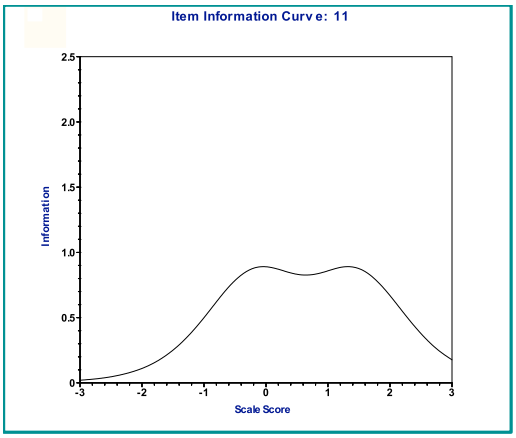


Category legends **Item: 10**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 11

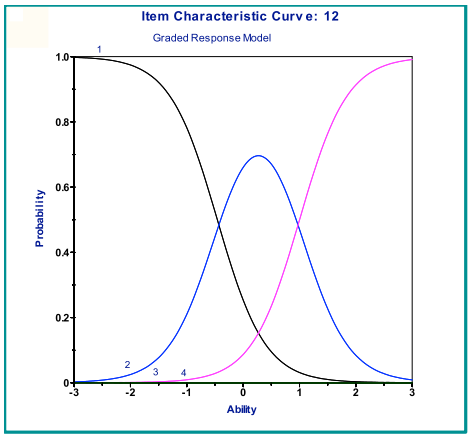


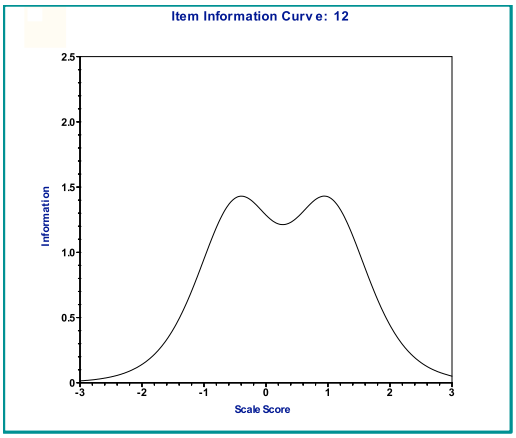


Category legends **Item: 11**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 12

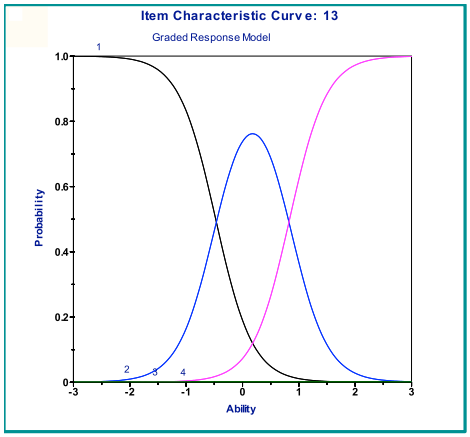


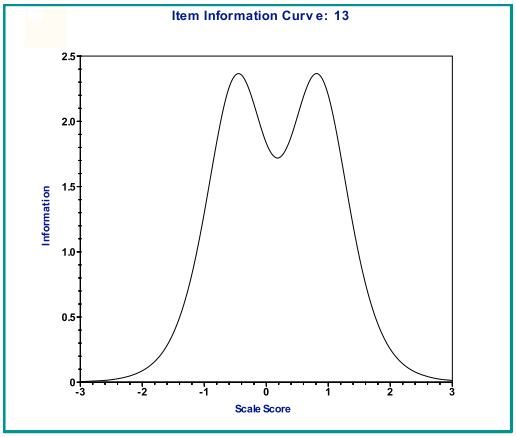


Category legends **Item: 12**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ÍTEM 13

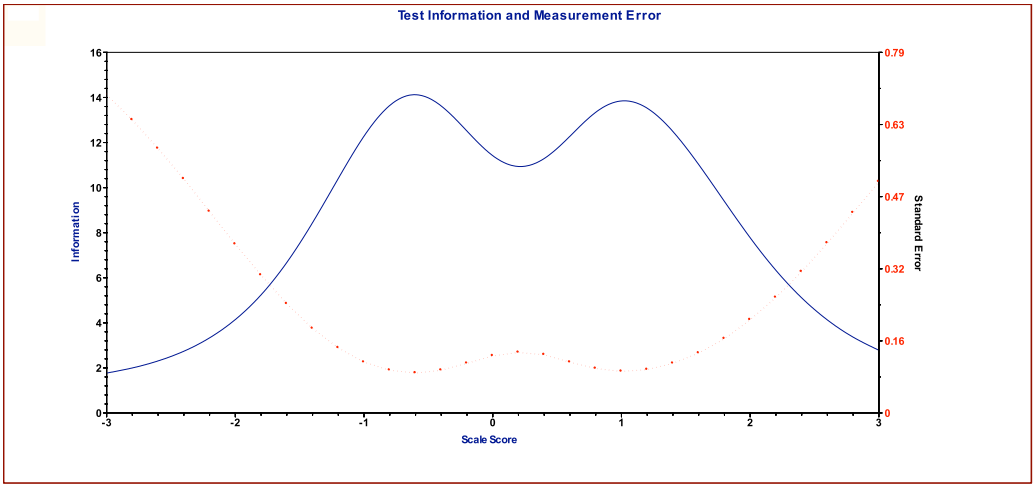




Category legends **Item: 13**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

CURVA CARACTERÍSTICA Y FUNCIÓN DE INFORMACIÓN DEL TEST



Test information curve: solid line **Standard error curve: dotted line**

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

8.3 Contraste de Levene calculado en el estudio sobre Experiencia

Prueba de homogeneidad de varianzas

	Estadístico de Levene	gl1	gl2	Sig.
Total Conducción Económica	**20,680**	**3**	**473**	**,000**
Total Seguridad	**2,116**	**3**	**473**	**,097**
Total Conducción Técnica	**10,601**	**3**	**473**	**,000**
Puntuación Total Test	**12,109**	**3**	**473**	**,000**

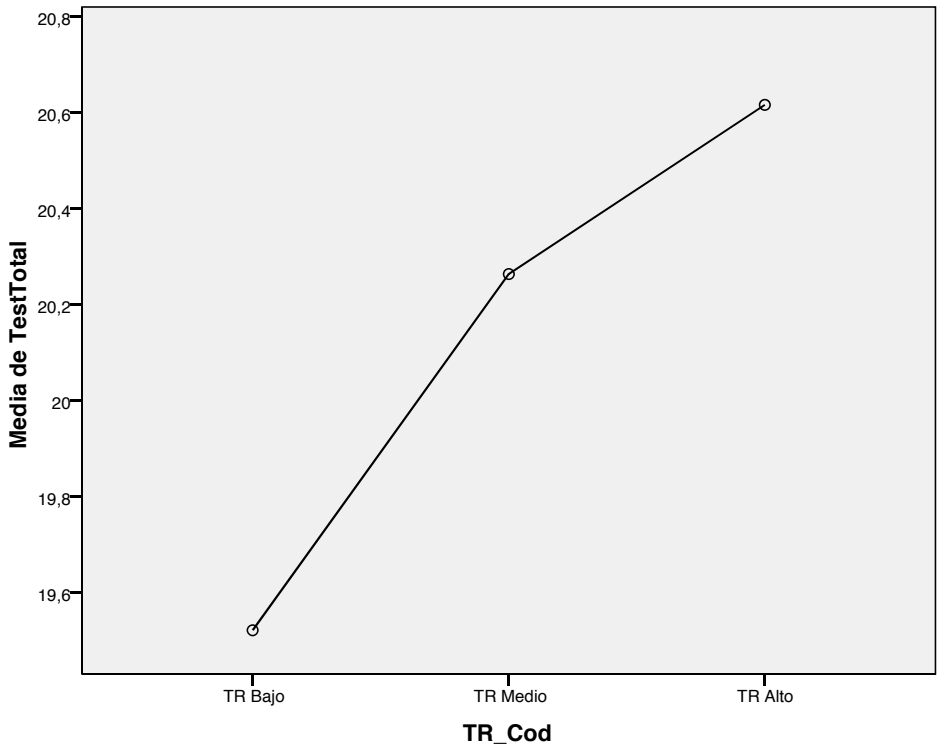
8.4 Resultados del las escalas del d2 en el estudio de Atención

Análisis de Varianza
Evaluación de conducción con Total Respuestas d2 (TR)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	4,912	2	2,456	,734	,483
	Intra-grupos	344,748	103	3,347		
	Total	349,660	105			
SegTot	Inter-grupos	2,594	2	1,297	,479	,621
	Intra-grupos	278,765	103	2,706		
	Total	281,358	105			
TecTot	Inter-grupos	,997	2	,499	,126	,882
	Intra-grupos	407,418	103	3,956		
	Total	408,415	105			
TestTotal	Inter-grupos	15,281	2	7,640	,334	,717
	Intra-grupos	2354,946	103	22,864		
	Total	2370,226	105			

Gráfico de las medias

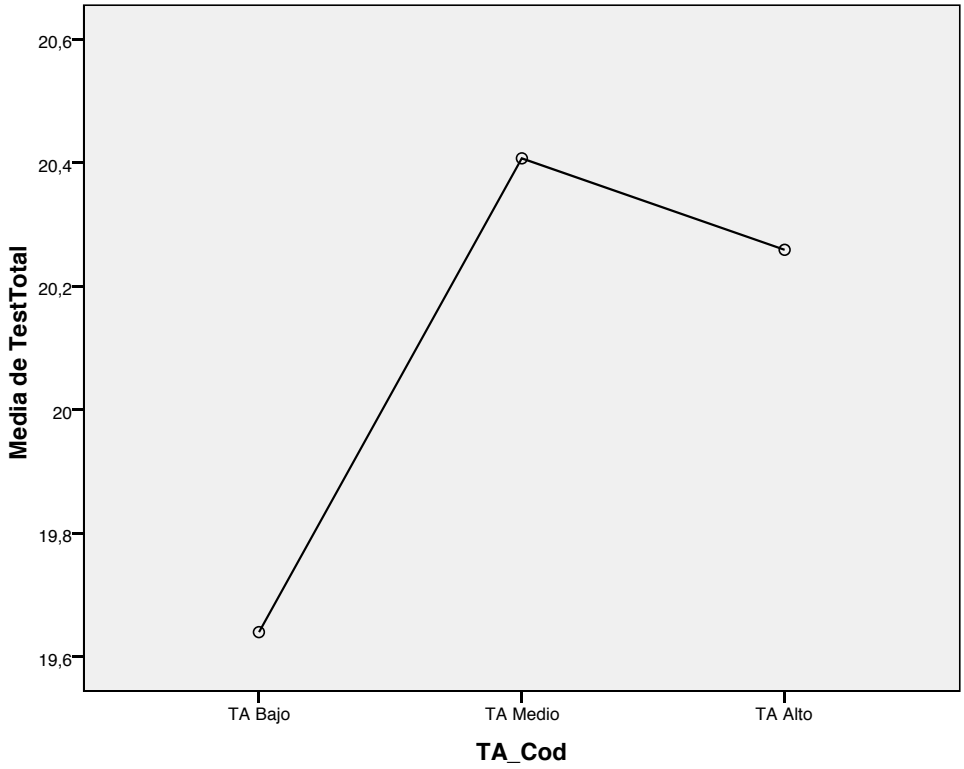


Análisis de Varianza
Evaluación de conducción con Total Aciertos d2 (TA)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	1,076	2	,538	,159	,853
	Intra-grupos	348,584	103	3,384		
	Total	349,660	105			
SegTot	Inter-grupos	3,030	2	1,515	,561	,573
	Intra-grupos	278,329	103	2,702		
	Total	281,358	105			
TecTot	Inter-grupos	,311	2	,155	,039	,962
	Intra-grupos	408,104	103	3,962		
	Total	408,415	105			
TestTotal	Inter-grupos	10,244	2	5,122	,224	,800
	Intra-grupos	2359,982	103	22,912		
	Total	2370,226	105			

Gráfico de las medias

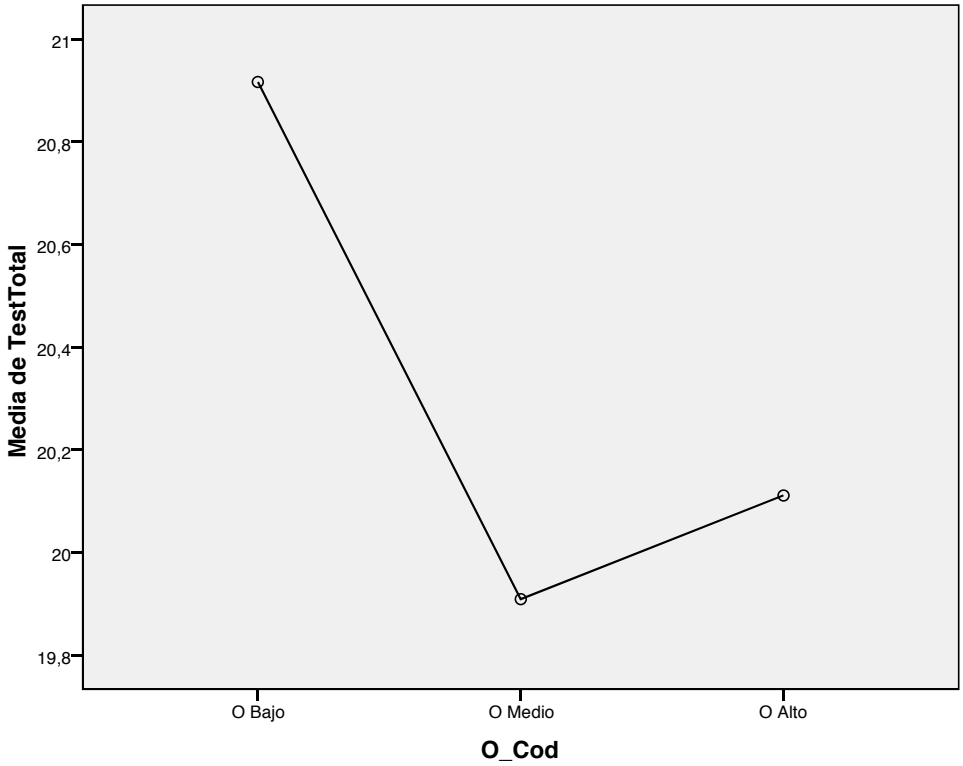


Análisis de Varianza
Evaluación de conducción con Omisiones d2 (O)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	,711	2	,355	,105	,901
	Intra-grupos	348,950	103	3,388		
	Total	349,660	105			
SegTot	Inter-grupos	9,233	2	4,617	1,747	,179
	Intra-grupos	272,125	103	2,642		
	Total	281,358	105			
TecTot	Inter-grupos	3,857	2	1,929	,491	,613
	Intra-grupos	404,558	103	3,928		
	Total	408,415	105			
TestTotal	Inter-grupos	17,181	2	8,590	,376	,688
	Intra-grupos	2353,045	103	22,845		
	Total	2370,226	105			

Gráfico de las medias

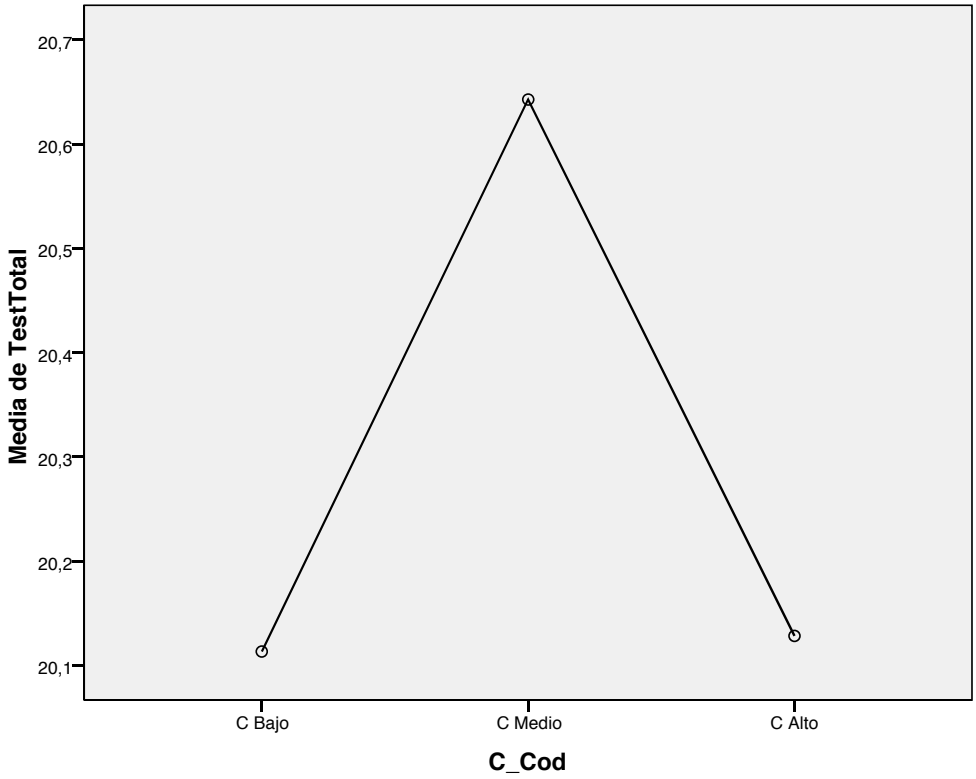


Análisis de Varianza
Evaluación de conducción con Comisiones d2 (C)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	**1,289**	**2**	**,644**	**,191**	**,827**
	Intra-grupos	**348,371**	**103**	**3,382**		
	Total	**349,660**	**105**			
SegTot	Inter-grupos	**,324**	**2**	**,162**	**,059**	**,942**
	Intra-grupos	**281,035**	**103**	**2,728**		
	Total	**281,358**	**105**			
TecTot	Inter-grupos	**2,708**	**2**	**1,354**	**,344**	**,710**
	Intra-grupos	**405,708**	**103**	**3,939**		
	Total	**408,415**	**105**			
TestTotal	Inter-grupos	**3,332**	**2**	**1,666**	**,073**	**,930**
	Intra-grupos	**2366,894**	**103**	**22,980**		
	Total	**2370,226**	**105**			

Gráfico de las medias

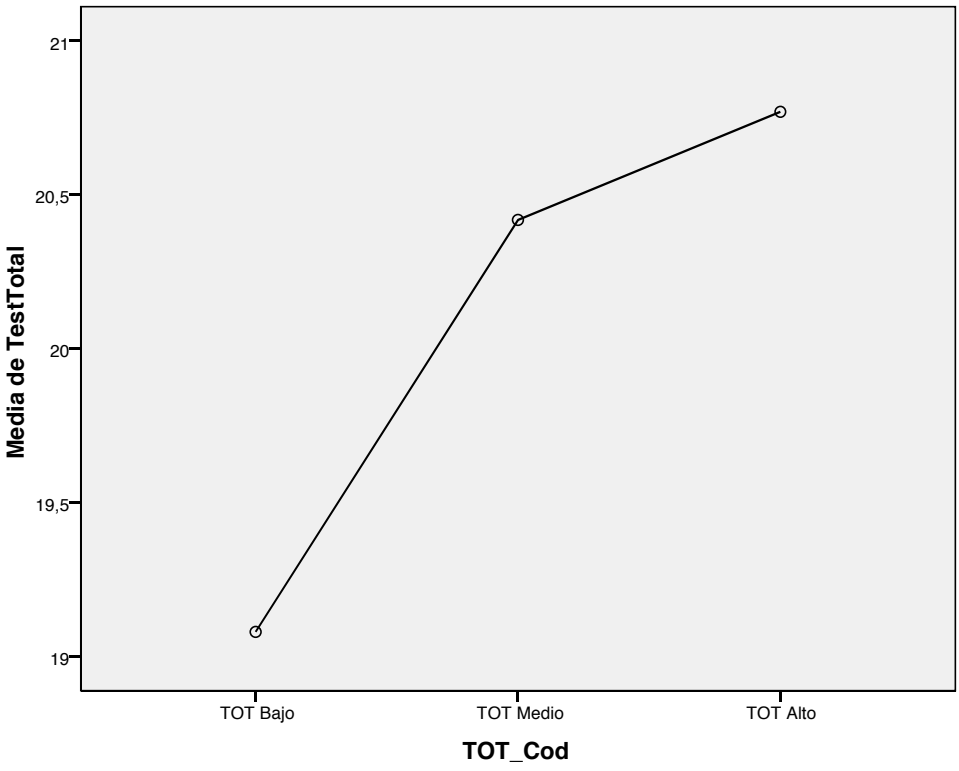


Análisis de Varianza
Evaluación de conducción con Total Test d2 (TOT)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	3,708	2	1,854	,552	,577
	Intra-grupos	345,952	103	3,359		
	Total	349,660	105			
SegTot	Inter-grupos	9,044	2	4,522	1,710	,186
	Intra-grupos	272,315	103	2,644		
	Total	281,358	105			
TecTot	Inter-grupos	3,591	2	1,796	,457	,635
	Intra-grupos	404,824	103	3,930		
	Total	408,415	105			
TestTotal	Inter-grupos	42,389	2	21,195	,938	,395
	Intra-grupos	2327,837	103	22,600		
	Total	2370,226	105			

Gráfico de las medias

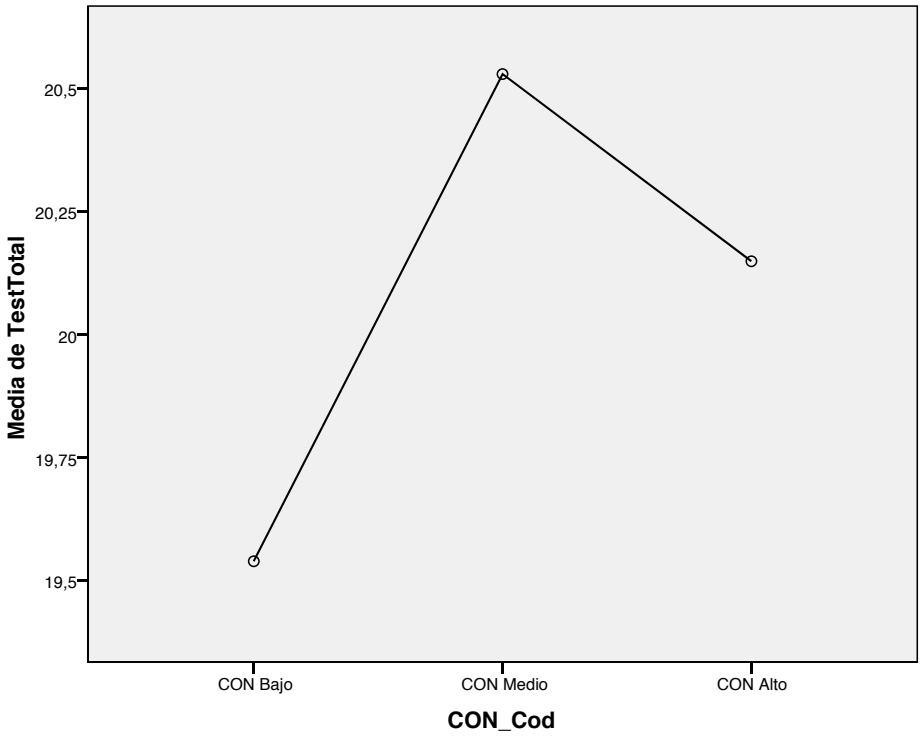


Análsis de Varianza
Evaluación de conducción con Concentración d2 (CON)

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
EcoTot	Inter-grupos	1,631	2	,816	,241	,786
	Intra-grupos	348,029	103	3,379		
	Total	349,660	105			
SegTot	Inter-grupos	4,442	2	2,221	,826	,441
	Intra-grupos	276,917	103	2,689		
	Total	281,358	105			
TecTot	Inter-grupos	,649	2	,324	,082	,921
	Intra-grupos	407,766	103	3,959		
	Total	408,415	105			
TestTotal	Inter-grupos	17,150	2	8,575	,375	,688
	Intra-grupos	2353,076	103	22,845		
	Total	2370,226	105			

Gráfico de las medias



8.5 Tabla de distribución de frecuencias de puntuaciones del NEO

Frecuencias Neuroticismo NEO

Estadísticos

NEO_N_Cod		
N	Válidos	**101**
	Perdidos	**3756**

NEO_N_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**32**	**,8**	**31,7**	**31,7**
	2,00	**44**	**1,1**	**43,6**	**75,2**
	3,00	**25**	**,6**	**24,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

Frecuencias Extraversión NEO

Estadísticos

NEO_E_Cod		
N	Válidos	**101**
	Perdidos	**3756**

NEO_E_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**25**	**,6**	**24,8**	**24,8**
	2,00	**57**	**1,5**	**56,4**	**81,2**
	3,00	**19**	**,5**	**18,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

Frecuencias Apertura NEO

Estadísticos

NEO_O_Cod		
N	Válidos	**101**
	Perdidos	**3756**

NEO_O_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**30**	**,8**	**29,7**	**29,7**
	2,00	**48**	**1,2**	**47,5**	**77,2**
	3,00	**23**	**,6**	**22,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

Frecuencias Amabilidad NEO

Estadísticos

NEO_A_Cod		
N	Válidos	**101**
	Perdidos	**3756**

NEO_A_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**35**	**,9**	**34,7**	**34,7**
	2,00	**48**	**1,2**	**47,5**	**82,2**
	3,00	**18**	**,5**	**17,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

Frecuencias Responsabilidad NEO

Estadísticos

NEO_C_Cod		
N	Válidos	**101**
	Perdidos	**3756**

NEO_C_Cod

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	**28**	**,7**	**27,7**	**27,7**
	2,00	**51**	**1,3**	**50,5**	**78,2**
	3,00	**22**	**,6**	**21,8**	**100,0**
	Total	**101**	**2,6**	**100,0**	
Perdidos	Sistema	**3756**	**97,4**		
Total		**3857**	**100,0**		

8.6 Análisis de varianza del estudio de personalidad

Factores inter-sujetos

		N
NEO_N_Cod	1,00	**32**
	2,00	**44**
	3,00	**25**
NEO_E_Cod	1,00	**25**
	2,00	**57**
	3,00	**19**
NEO_O_Cod	1,00	**30**
	2,00	**48**
	3,00	**23**
NEO_A_Cod	1,00	**35**
	2,00	**48**
	3,00	**18**
NEO_C_Cod	1,00	**28**
	2,00	**51**
	3,00	**22**

Pruebas de los efectos inter-sujetos

Variable dependiente:Puntuación Total Test

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	**1092,367**[a]	**47**	**23,242**	**,674**	**,914**
Intersección	**22992,372**	**1**	**22992,372**	**667,145**	**,000**
NEO_N_Cod	**14,506**	**2**	**7,253**	**,210**	**,811**
NEO_E_Cod	**15,840**	**2**	**7,920**	**,230**	**,795**
NEO_O_Cod	**54,860**	**2**	**27,430**	**,796**	**,456**
NEO_A_Cod	**73,051**	**2**	**36,525**	**1,060**	**,354**
NEO_C_Cod	**11,227**	**2**	**5,613**	**,163**	**,850**
NEO_N_Cod * NEO_E_Cod	**,000**	**0**	**.**	**.**	**.**
NEO_N_Cod * NEO_O_Cod	**97,502**	**4**	**24,375**	**,707**	**,591**
NEO_N_Cod * NEO_A_Cod	**27,034**	**2**	**13,517**	**,392**	**,678**
NEO_N_Cod * NEO_C_Cod	**10,889**	**1**	**10,889**	**,316**	**,576**
NEO_E_Cod * NEO_O_Cod	**152,442**	**1**	**152,442**	**4,423**	**,040**
NEO_E_Cod * NEO_A_Cod	**42,579**	**1**	**42,579**	**1,235**	**,271**
NEO_E_Cod * NEO_C_Cod	**48,039**	**1**	**48,039**	**1,394**	**,243**
NEO_O_Cod * NEO_A_Cod	**,000**	**1**	**,000**	**,000**	**1,000**
NEO_O_Cod * NEO_C_Cod	**126,298**	**1**	**126,298**	**3,665**	**,061**
NEO_A_Cod * NEO_C_Cod	**135,223**	**3**	**45,074**	**1,308**	**,282**
NEO_N_Cod * NEO_E_Cod * NEO_O_Cod	**,000**	**0**	**.**	**.**	**.**
NEO_N_Cod * NEO_E_Cod * NEO_A_Cod	**,000**	**0**	**.**	**.**	**.**
NEO_N_Cod * NEO_E_Cod * NEO_C_Cod	**,000**	**0**	**.**	**.**	**.**

a. R cuadrado = ,374 (R cuadrado corregida = -,181)