Consejo General de la Psicología ESPAÑA

Psicothema (2022) 34(4) 603-609

Psicothema

https://www.psicothema.com/es • ISSN 0214–7823

Colegio Oficial de Psicólogos del Principado de Asturias

Psicothema

Methodology

Driving Offences and Emotion Regulation: A Psychometric Analysis of the Emotion Regulation Questionnaire (ERQ)

Jennifer Pérez-Sánchez[1], Ana R. Delgado[1] and Gerardo Prieto[1]

1 Universidad de Salamanca.

ARTICLE INFO

Received: March 09, 2022
Accepted: May 02, 2022

Keywords:
Emotion regulation
ERQ
Rasch Model
Rating Scale Model
Traffic psychology

ABSTRACT

Background: Little is known about the role of emotion regulation in driving. This study analyzed the psychometric properties of one of the most commonly-used tests for the evaluation of emotion regulation in a drivers' sample. **Methods:** A total of 318 male drivers (M age = 41.6 years, SD = 11.1, age range 20-69 years, half with road traffic offences and half, matched controls), participated in the study by filling out the Emotion Regulation Questionnaire (ERQ). Data analysis was carried out using the Rasch Rating Scale Model (RSM). **Results:** The performance of the response categories was inadequate. After collapsing the seven original response categories into three, the categories were functional, the unidimensionality requirement was met, and data-model fit was adequate. Both person reliability and item reliability were adequate for the two subscales of ERQ. There was a significant difference between the two groups in the cognitive reappraisal subscale. **Conclusions:** Given the adequate psychometric properties of the ERQ, it is suggested that the role of cognitive reappraisal in driving contexts be explored in depth.

Infracciones de Tráfico y Regulación Emocional: Análisis Psicométrico del ERQ

Palabras clave:
Regulación emocional
ERQ
Modelo de Rasch
Modelo de escalas de calificación
Psicología del tráfico

RESUMEN

Antecedentes: No es mucho lo que se conoce sobre el papel de la regulación emocional en la conducción. Analizamos las propiedades psicométricas de uno de los test más utilizados para la evaluación de la regulación emocional en una muestra de conductores. **Métodos:** un total de 318 conductores varones (M = 41.6 años, DT = 11.1, rango de edad 20-69 años; la mitad con infracciones de tráfico y la mitad restante, controles emparejados), cumplimentaron el cuestionario de regulación de emociones (ERQ en inglés). El análisis de los datos se llevó a cabo mediante el modelo de Rasch para escalas de calificación. **Resultados:** El funcionamiento de las categorías de respuesta fue inadecuado. Tras reducir las siete categorías de respuesta originales a tres, las categorías resultaron funcionales, se cumplía el requisito de unidimensionalidad y el ajuste de los datos con el modelo fue adecuado. Tanto la fiabilidad de las personas como la de los ítems fueron adecuadas en las dos subescalas del ERQ. Se halló una diferencia significativa entre los grupos de estudio en la subescala de revaluación cognitiva. **Conclusiones:** Dadas las adecuadas propiedades psicométricas del ERQ, se sugiere explorar en profundidad el papel de la revaluación cognitiva en contextos de conducción.

Cite as: Pérez-Sánchez, J., Delgado, A. R., & Prieto, G. (2022). Driving Offences and Emotion Regulation: A Psychometric Analysis of the Emotion Regulation Questionnaire (ERQ). *Psicothema, 34*(4), 603-609. https://doi.org/10.7334/psicothema2022.120
Corresponding author: Jennifer Pérez-Sánchez, jeps@usal.es

Driving involves keeping control over of the vehicle and responding adaptatively to traffic situations (Trógolo et al., 2014). To meet traffic standards, this activity requires cognitive resources such as reaction capacity and decision-making (Chan & Singhal, 2013). The cornerstones of driving are the road, the vehicle and the human (Zhang et al., 2013). The latter is the main factor that can cause but also prevent road accidents. Road safety agencies world-wide seek to reduce risky and reckless driving behavior (Carey et al., 2013). Traffic advertising campaigns usually employ shocking content even though effectiveness might not be as high as expected (Eherenfreund-Hager et al., 2017; Fischer et al., 2011). Another progressive system to promote good coexistence on roads is the system that consists of points and monetary fines. There are different penalties (change of points, fine payment, prison sentences) that can put peoples' driving licenses at risk. Many countries in Europe use this system. In 2006, Spain adopted what they called "permiso de conducir por puntos" (González et al., 2008; Ministerio del Interior, 2005). Drivers get up to 15 intangible points that can be removed. Lack of points or serious infractions disqualify drivers.

This system allows drivers to voluntarily recover some of their lost points. In this case, they must attend specific courses to recover their driving license. Spanish driving schools offer two types of courses: *partial course*, where it is possible to recover up to 6 points and *total course*, where the aim is to recover the driving license.

Among psychological factors, dysfunctions in attention or decision-making are plausible causes of errors during driving. Nevertheless, there are other relevant variables such as emotions which have not been studied to the same degree (Fang et al., 2020; Taubman-Ben-Ari, 2012). Emotion Regulation (ER) refers to the way people influence the occurrence, intensity, duration and expression of emotions (Gross, 2015). The psychological variable ER has witnessed the biggest increase in interest in recent decades. It has been widely studied in clinical and educational fields. However, it is considered less in contexts such as traffic where research is focused on relationships between driving styles and ER (Navon-Eyal & Taubman-Ben-Ari, 2019).

The process model (Gross, 2015) is the usual model to explain ER. It combines generation and regulation of emotions. An emotion is generated through four phases (situation, attention, valuation and emotional response). Five groups of ER strategies can intervene in each phase (situation selection, situation modification, attentional deployment, cognitive change and modulation response). Recently, a further four superior phases have been added: identification, selection, implementation and monitoring (McRae & Gross, 2020).

So far, cognitive reappraisal and expressive suppression have been the two most studied strategies. In the process model, the former belongs to the cognitive change group whilst the latter is the prototype for the modulation response group. Ever since the process model was proposed, strategies can be classified as located before (antecedent-focused) or after the emotional response takes place (consequent-focused) (Gross, 1998).

Over the years, cognitive reappraisal and expressive suppression have been assessed by means of various instruments (McRae & Gross, 2020), mainly self-reports (Pérez-Sánchez et al., 2020a). *Emotion Regulation Questionnaire* (ERQ), developed by Gross and John (2003), is one of the most used tests in the evaluation of the ER (Pérez-Sánchez et al., 2020a). It is a 10-item test that consists of two subscales: cognitive reappraisal (6 items) and expressive suppression (4 items). Response categories are 7-Likert-format, from 1 (strongly disagree) to 7 (strongly agree). There is not a total score. The higher the score in a subscale, the more that strategy is used to regulate emotions. For this original version, score internal consistency coefficients were adequate for both cognitive reappraisal (α = .79) and expressive suppression scores (α = .73). As to validity, there was evidence from correlations between the scores of the ERQ and criteria such as tests that assessed coping strategies, emotional intelligence, personality traits, affective and social functioning, and long-term well-being (Pérez-Sánchez et al., 2020b).

Classical Test Theory (CTT) is a handy approach that dominated psychometrics in the past century. Nevertheless, one of the main disadvantages of using CTT is that scores lack measurement invariance (Muñiz, 2010). The Rating Scale Model (RSM, Andrich, 1978) is an extension of the Rasch Model (RM) for polytomous data:

$$(1) \qquad \ln (P_{nik}/(P_{ni(k-1)})) = B_n - D_i - F_k$$

where:

P_{nik} is the probability that person n on encountering item i would be observed (or would respond) in category k,

B_n is the level of person n,

D_i is the location of item i,

F_k is a rating scale threshold defined as the location corresponding to the equal probability of observing adjacent categories k-1 and k.

Both RM and RSM share desirable properties of invariant measurement: a) item-invariant measurement of persons, b) person-invariant calibration of test items, c) non-crossing person response functions, d) non-crossing item response functions, and e) items and persons parameters must be located on a single latent variable (Engelhard & Wang, 2021). Moreover, the representation of items and persons on the same scale is possible, provided that invariant measurement is met (Prieto & Delgado, 2003).

The number and performance of response categories is highly relevant for the utility and validity of scores (Simms et al., 2019). Therefore, the functionality of rating scales should be tested empirically on the observed data set. According to Linacre (2002), the criteria are as follows: a) high frequencies in each category, at least 10 observed counts, b) acceptable distribution of observations in the categories: uniform, unimodal peaking in central or extreme categories, or bimodal distributions peaking in extreme categories), c) the average person measures must go monotonically up the rating scale, d) *infit* and *outfit* mean squares statistics in each category have to be less than 2.00, and e) thresholds between categories have to be ordered monotonically. If response categories do not work properly, adjacent categories have to be collapsed (Cavanagh & Fisher, 2018).

The overarching goal of this study was to test the psychometric properties of ERQ scores from a Spanish drivers' sample using the RSM. To do so, we examined the functionality of response categories, unidimensionality, data-fit and conjoint measurement. We also obtained evidence of validity based on the absence of Differential Item Functioning (DIF) associated with groups with different traffic sanction history and with different age. We also obtained the evidence of validity based on the differences between the ERQ mean scores of these groups of drivers.

Method

Participants

A total of 318 male drivers (*M* age = 41.6 years, *SD* = 11.1, age range 20-69 years). Most participants had completed high or middle school (69%) and drove virtually every day (89%) while they had a valid driving license. The sample only included males, as there were not many females with road traffic offences.

The drivers that had lost points or their driving license were included in the group of drivers with road traffic offences (n=159). A comparison group (n=159) was formed with non-offender volunteers matched in driving frequency, educational level and age (+ 3 years old, so that there was no a significant difference in age between both groups, t = -.22, p = .859, d = -.02).

Instruments

Sociodemographic and driving frequency questions were used to record variables such as age, gender, education and driving frequency of weekly driving.

The authorized Spanish translation and adaptation of the original version of the Emotion Regulation Questionnaire (ERQ; Gross, 2020). A 10-item scale assessing dispositional tendencies to use two ER strategies: cognitive reappraisal (6 items, e.g., *"I control my emotions by changing the way I think about the situation I'm in"*) and expressive suppression (4 items, e.g., *"I keep my emotions to myself"*). Participants were asked to what degree each of the statements applied to them, responding on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree).

Procedure

The study was approved by the Dirección General de Tráfico (DGT), two driving schools in Salamanca (Spain), and the bioethics committee of the University of Salamanca. The participants were recruited by means of convenience sampling. All potential participants received a brief explanation about the aims of the research and confidentiality was ensured. The total time required to complete the questionnaire for each participant was 15-20 minutes.

Data Analysis

Responses to cognitive reappraisal subscale and expressive suppression subscale were separately analyzed according to RSM (Andrich, 1978). Data analysis were performed using the computer program Winsteps 4.7.0 (Linacre, 2021).

The optimal number and performance of response categories was examined empirically by means of Linacre (2002) criteria.

Unidimensionality was tested by means of the Principal Component Analysis (PCA) of the residuals. According to Linacre (2022) guidelines, the typical range for the variance explained by the Rasch measure is 40%-50%, and the eigenvalue of the first contrast should be less than 3.0. The assumption of local independence was assessed with Yen's Q3 test. High positive correlations between residuals would be clearly indicative of local dependence.

Data-model fit was assessed by *infit* (an information-weighted statistic) and outfit (an unweighted statistic). Infit/outfit values over 2 degrade measurement (Linacre, 2022).

Reliability statistics, such as Item Separation Reliability (ISR) and Person Separation Reliability (PSR), assess the accuracy of the item and person estimations by indicating the proportion of the observed variance that is reproducible from Rasch model. ISR and PSR values over .70 are recommended to achieve a suitable measure.

DIF is a systematic error that can affect the *validity* of the scores. Its presence reveals the inclusion of construct-irrelevant variance in scores. The presence of DIF is manifested when the probability of a response to an item is unequal for two groups of individuals, conventionally called the reference group and the focal group, once the individuals in each group have been matched on the variable of interest (Prieto & Nieto, 2014). The empirical evidence of presence of DIF is confirmed when the difference between location item parameters for each group is equal or over |.64| logits and statistically significant, considering Welch's t-test and Bonferroni-corrected *alpha* levels (Linacre, 2022).

Finally, evidence of criterion validity was obtained by means of *impact* analysis, operationalized by the differences between the means on the scales of groups related to the criterion of interest, in this case the history of traffic sanctions. *Welch's t-test* and *Cohen's d* were calculated to assess statistical significance and magnitude of impact, respectively.

Results

Cognitive Reappraisal

The psychometric quality of the seven original response categories for the cognitive reappraisal subscale was analyzed. Table 1 shows that Linacre (2002) criteria were not met for the seven original categories. The thresholds between the categories are not arranged in a monotonic form.

Table 1.
Cognitive Reappraisal (ERQ): Analysis of the Seven Original Categories.

Category	Count	Mean	Infit	Outfit	Threshold
1	148 (8%)	-.53	1.05	1.29	-
2	167 (9%)	-.32	1.00	1.17	-.55
3	116 (6%)	-.11	.97	.99	.15
4	312 (16%)	.06	.79	.72	-1.01
5	321 (17%)	.28	.74	.68	.16
6	543 (28%)	.59	.95	1.01	-.10
7	301 (16%)	.88	1.29	1.16	1.35

Count: number of counts observed in each category; Mean: Mean of the differences in each category between the person and item parameters; Outfit: Fit statistics of categories; Threshold: value between adjacent categories.

Therefore, it was necessary to collapse successive categories into three categories (two at the bottom, three in the middle and two at the top). The performance of the three collapsed categories was adequate (Table 2) and the rest of the analyses were performed considering these three categories (low, medium and high agreement).

The percentage of variance explained by cognitive reappraisal measures was 35.7% and the eigenvalue of the first contrast of the unexplained variance was 1.55. The percentage of explained variance was slightly below the typical range, but the first contrast indicated that there was enough evidence for essential unidimensionality. As to local independence, there were no positive correlations between the residuals, which ranged from -.02 to -.30. Table 3 shows that

there was good data-model fit for items, mean *infit* was 1.0 (SD = 0.2) and mean *outfit* was 1.0 (SD = 0.2). No item showed overfitting or misfit over 1.5. For persons, mean *infit* was 1.0 (SD = 0.7) and mean *outfit* was 1.0 (SD = 0.7). Thirty-three participants (11.4%) showed infit over 2, and twenty-eight (9.7%) showed *outfit* over 2.

There were no extreme scores for items. Twenty-eight persons got extreme scores: twenty-six got the maximum (4.02 *logits*) and two the minimum (-4.03 *logits*).

Person measures are located at the left of the Wright map while the right side shows item locations (Figure 1). Average person measure was .87 *logits* (SD = 1.5). Average item locations are conventionally 0.00 *logits*. Item separation reliability (*ISR*=.93) was good enough. Although, the parameters of the person's level were estimated with a slightly low degree of accuracy (*PSR*=.65).

No item showed group- or age-related DIF (Table 4). The mean performance of the group of matched controls was higher than that of the drivers with road traffic offences. Although statistically significant (*Welch-t* (314) = -2.19, p = .029), the magnitude of the difference between groups was small (d = .25).

Table 2.
Cognitive Reappraisal (ERQ): Analysis of the Three Collapsed Categories.

Category	Count	Mean	Infit	Outfit	Threshold
1	315 (17%)	-.68	1.01	1.05	-
2	749 (39%)	.45	.88	.89	-1.01
3	844 (44%)	1.39	1.05	1.09	1.01

Count: number of counts observed in each category; Mean: Mean of the differences in each category between the person and item parameters; Threshold: value between adjacent categories.

Table 3.
Cognitive Reappraisal (ERQ): Item Statistics.

Item	Di	SE	Infit	Outfit	Item-Measure
1	.41	.10	.93	.93	.67
3	-.16	.10	.97	.93	.64
5	-.65	.11	1.39	1.41	.49
7	.26	.10	1.01	1.04	.64
8	.35	.10	.84	.94	.65
10	-.20	.10	.91	.88	.65

D_i: item location parameter; SE: Standard Error; Item-Measure: Item-Measure correlation.

Table 4.
Cognitive Reappraisal (ERQ): Group-and-Age-Related DIF.

Group-related DIF:						
D a	**D b**	**DIF Contrast**	**Joint S.E**	***t-Welch***	***P***	**Item**
0.58	0.23	0.35	0.20	1.78	.076	1
-0.29	-0.04	-0.25	0.20	-1.22	.222	3
-0.72	-0.59	-0.12	0.22	-0.58	.564	5
0.37	0.15	0.22	0.20	1.11	.267	7
0.43	0.27	0.16	0.20	0.80	.424	8
-0.41	0.01	-0.42	0.20	-2.03	.042	10
Age-related DIF:						
D c	**D d**	**DIF Contrast**	**Joint S.E**	***t-Welch***	***P***	**Item**
0.35	0.47	-0.12	0.20	-0.61	.542	1
-0.42	0.09	-0.52	0.21	-2.52	.012	3
-0.54	-0.78	0.25	0.22	1.15	.252	5
0.29	0.23	0.06	0.20	.29	.768	7
0.52	0.17	0.35	0.20	1.77	.078	8
-0.20	-0.20	0.00	0.20	0.00	1.00	10

D a: Item location parameter in the group of drivers with road traffic offences; D b: Item location parameter in the comparison group. D c: Item location parameter in the group of participants below 41 years old; D d: Item location parameter in the group of participants above 41 years old; DIF Contrast: difference between both parameters (Da-Db or Dc-Dd). Joint S.E: the standard error of the DIF Contrast.

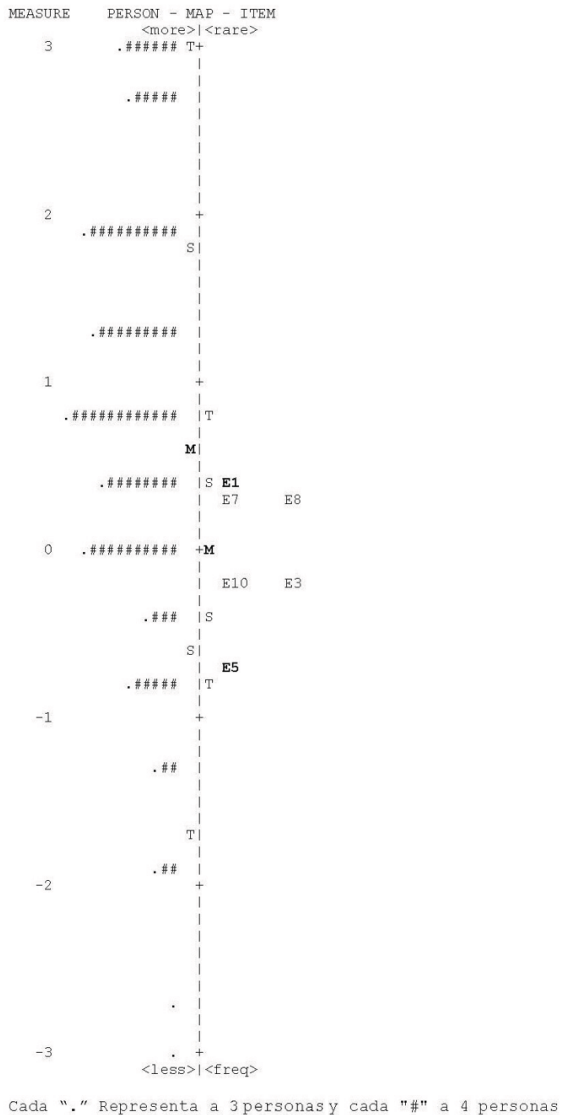

Figure 1.
Cognitive Reappraisal (ERQ): Conjoint Measurement.

Expressive Suppression

Seven original response categories showed inadequate functioning according to criteria proposed by Linacre (2002). In Table 5 it can be seen that the thresholds between the response categories were disordered.

Therefore, it was necessary to collapse successive categories into three categories (two at the bottom, three in the middle and two at the top). The performance of the three collapsed categories were adequate (Table 6) and the rest of the analyses were performed considering these three categories (low, medium and high agreement).

"The percentage of variance explained by expressive suppression measures was 40.1% and the eigenvalue of the first contrast of the unexplained variance was 1.39. Therefore, the empirical conditions to assume unidimensionality on this scale were met. As to local independence, there were no positive correlations between the residuals, which ranged from -.24 to -.38. Table 7 shows that there was good data-model fit for items, the mean *infit* was 1.0 (SD = 0.1)

and the mean *outfit* was .99 (SD = 0.1). No item showed overfitting or misfit greater than 1.5. As for data-model fit for persons, it was good enough: the mean *infit* was 1.0 (SD=0.7) and the mean *outfit* was .99 (SD = 0.8). Only 10.6% (29 participants) showed *infit* over 2; *outfit* was greater than 2 for 9.2% of the sample (24 participants). There were no extreme scores for items. Forty-five persons got extreme scores: twenty-two got the maximum (2.61 *logits*), and twenty-three the minimum (-2.61 *logits*).

Figure 2 shows the conjoint scaling. Average person aptitude was -.02 *logits* (SD = 1.5). Item separation reliability (*ISR*=.94) was good enough. Although, the parameters of the person's level were estimated with a slightly low degree of accuracy (*PSR*=.68).

No item showed group- or age-related DIF (Table 8). The mean performance of the group of matched controls was located in the same place on the scale as the group with road traffic offences. Therefore, statistically significant group differences (impact) in Rasch measures were not found, *Welch-t* (315) = -.37, p = .709, d = .04.

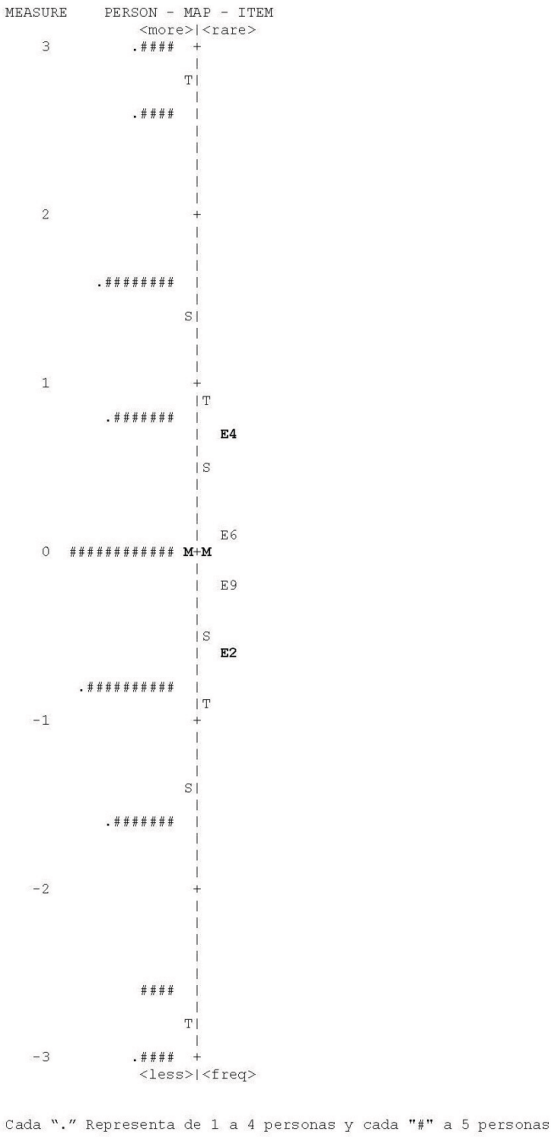

Figure 2.
Expressive Suppression (ERQ): Conjoint Measurement.

Table 5.
Expressive Suppression (ERQ): Analysis of the Seven Original Categories.

Category	Count	Mean	Infit	Outfit	Threshold
1	166 (13%)	-.78	1.42	1.46	-
2	196 (15%)	-.71	.69	.65	-1.15
3	163 (13%)	-.32	.90	.86	-.27
4	186 (15%)	-.06	.82	.74	-.33
5	207 (16%)	.18	.94	.82	-.04
6	225 (18%)	.66	.85	.89	.29
7	129 (10%)	1.00	1.36	1.22	1.51

Count: number of counts observed in each category; Mean: Mean of the differences in each category between the person and item parameters; Outfit: Fit statistics of categories; Threshold: value between adjacent categories.

Table 6.
Expressive Suppression (ERQ): Analysis of the Three Collapsed Categories.

Category	Count	Mean	Infit	Outfit	Threshold
1	362 (28%)	-1.33	1.03	1.05	-
2	556 (44%)	-.01	.92	.86	-1.41
3	354 (28%)	1.34	1.03	1.03	1.41

Count: number of counts observed in each category; Mean: Mean of the differences in each category between the person and item parameters; Threshold: value between adjacent categories.

Table 7.
Expressive Suppression (ERQ): Item Statistics.

Item	D_i	*SE*	*Infit*	*Outfit*	Item-Measure
2	-.57	.11	1.11	1.10	.71
4	.67	.11	1.12	1.07	.72
6	.15	.11	.84	.84	.77
9	-.25	.11	.94	.95	.73

D_i: item location parameter; SE: Standard Error; Item-Measure: Item-Measure correlation.

Table 8.
Expressive Suppression (ERQ): Group-and-Age-Related DIF.

Group-related DIF:						
D a	**D b**	**DIF Contrast**	**Joint S.E**	***t-Welch***	***p***	**Item**
-0.41	-0.73	0.32	0.22	1.41	.159	2
0.67	0.69	-0.02	0.23	-0.09	.929	4
0.23	0.07	0.16	0.22	0.71	.475	6
-0.46	-0.03	-0.43	0.22	-1.95	.052	9
Age-related DIF:						
D c	**D d**	**DIF Contrast**	**Joint S.E**	***t-Welch***	***p***	**Item**
-0.85	-0.28	-0.57	0.22	-2.53	.012	2
0.80	0.55	0.25	0.23	1.13	.261	4
0.32	-0.03	0.35	0.22	1.56	.119	6
-0.25	-0.25	0.00	0.22	0.00	1.00	9

D a: Item location parameter in the group of drivers with road traffic offences; D b: Item location parameter in the comparison group. D c: Item location parameter in the group of participants below 41 years old; D d: Item location parameter in the group of participants above 41 years old; DIF Contrast: difference between both parameters (Da-Db or Dc-Dd). Joint S.E: the standard error of the DIF Contrast.

Discussion

The performance of the seven original response categories of ERQ for both (cognitive reappraisal and expressive suppression) subscales was inadequate. Using seven categories could create confusion for the participants as to the definition of the variable and the meaning of the scores. Collapsing these categories into three, ranging from 1 (disagree) to 3 (agree), served to meet the criteria proposed by Linacre (2002).

The unidimensionality requirement was also met, there was no local dependence, and data-model fit was good enough for both items and persons. Therefore, it was possible to measure conjointly items and persons on the same *logit* scale. The most severe items were of positive or neutral semantic content which is consistent with the hedonist perspective: positive or neutral emotions are less likely to be regulated than negative emotions (Nigg, 2017).

Since there was no presence group-related DIF (offenders vs. comparison) or age-related DIF either for the cognitive reappraisal scale or for the expressive suppression one, it was possible to test the difference between groups (impact). We found a statistically significant difference between mean scores of the offender group and the comparison group in the cognitive reappraisal subscale, which is consistent with previous studies: reappraisal is an adaptive ER strategy that has been associated with careful driving styles (Holman & Popusoi, 2020).

Notwithstanding the above-mentioned implications, the present research has limitations. Mainly, ERQ items do not define situations narrowly related to the driving context. Consequently, ERQ items may not properly capture the complex nature of ER in the driving field. Secondly, the variability of the items is low, so it would be convenient to include items at the extremes of the severity continuum. The inclusion of this type of items would contribute to increase the precision of the measurements of individuals. In addition, the formulation of items specific to the traffic context will contribute to improve the content validity of emotional regulation tests in that domain.

Finally, it would be interesting to carry on studies with other samples, such as those of females (even though there are few offenders among them) or other-culture samples. In this study, the gender variable was controlled by elimination as few women have committed road traffic offences and attend driving school courses in Spain.

In conclusion, the present study allows us to assess the quality of the ERQ scores from a Spanish drivers' sample, analyzed with a psychometric advanced model: the Rasch RSM. Our findings indicate that the ERQ scores have basically adequate metric properties, and that the cognitive reappraisal strategy might be useful in driving contexts. However, it should be noted that there is not a specific test of ER in driving, so future investigations could adapt the ERQ items to the traffic field.

Acknowledgements

This study is part of the Ph.D. dissertation submitted to the University of Salamanca by the first author, and was carried out under the supervision of the second and the third authors. The study was supported by a grant from the University of Salamanca and Santander Bank. We also express our gratitude to the two driving schools for their support.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573. https://doi.org/10.1007/BF02293814

Carey, R. N., McDermott, D. T., & Sarma, K. M. (2013). The impact of threat appeals on fear arousal and driver behavior: a meta-analysis of experimental research 1990-2011. *PloS ONE, 8*(5), Article e62821. https://doi.org/10.1371/journal.pone.0062821

Cavanagh, R. F., & Fisher, W. P. Jr. (2018). The science in human science research: The case for Rasch measurement in learning environment research. *Journal of Physics: Conference Series, 1065*(7), Article 072010. https://doi.org/10.1088/1742-6596/1065/7/072010

Chan, M., & Singhal, A. (2013). The emotional side of cognitive distraction: Implications for road safety. *Accident, Analysis & Prevention, 50*, 147-154. https://doi.org/10.1016/j.aap.2012.04.004

Eherenfreund-Hager, A., Taubman-Ben-Ari, O., Toledo, T., & Farah, H. (2017). The effect of positive and negative emotions on young drivers: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 49*, 236-243. https://doi.org/10.1016/j.trf.2017.07.002

Engelhard, G. Jr., & Wang, J. (2021). *Rasch Models for Solving Measurement Problems (Quantitative Applications in the Social Sciences)*. SAGE Publications.

Fang, C., Zhang, Y., Zhang, M., & Fang, Q. (2020). P300 measures and drive-related risks: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health, 17*(15), Article 5266. https://doi.org/10.3390/ijerph17155266

Fischer, P., Greitemeyer, T., Kastenmüller, A., Vogrincic, C., & Sauer, A. (2011). The effects of risk-glorifying media exposure on risk-positive cognitions, emotions, and behaviors: a meta-analytic review. *Psychological Bulletin, 137*(3), 367-390. https://doi.org/10.1037/a0022267

González, L. M., Ruiz, J. R., & Gil, F. T. (2008). Influencia del permiso de conducción por puntos en el comportamiento al volante: Percepción de los conductores [Influence of the penalty point system on driving behaviour]. *Psicothema, 20*(4), 652-658.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74*(1), 224-237. https://doi.org/10.1037/0022-3514.74.1.224

Gross, J. J. (2015). Emotion regulation: current status and future prospects. *Psychological Inquiry, 26*(1), 1-26. https://doi.org/10.1080/1047840X.2014.940781

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well- being. *Journal of Personality and Social Psychology, 85*(2), 348-362. https://doi.org/10.1037/0022-3514.85.2.348

Gross, J. J. (2020, November 9). *The emotion regulation questionnaire*. https://spl.stanford.edu/resources

Holman A. C., & Popușoi, S. A. (2020). How you deal with your emotions is how you drive. emotion regulation strategies, traffic offenses, and the mediating role of driving styles. *Sustainability, 12*(12), Article 4929. https://doi.org/10.3390/su12124929

Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2021). *Winsteps®* (Version 4.7.0) [Computer Software]. https://www.winsteps.com/index.htm

Linacre, J. M. (2022, January 1). *Winsteps Help for Rasch Analysis*. https://www.winsteps.com/winman/index.htm

McRae, K., & Gross, J. J. (2020). Emotion Regulation. *American Psychological Association, 20*(1), 1-9. https://doi.org/10.1037/emo0000703

Ministerio del Interior (2005, 28 de julio). *ORDEN INT/2596: Por la que se regulan los cursos de sensibilización y reeducación vial para los titulares de un permiso o licencia de conducción [Order INT/2596/2005: Regulating awareness and re-education courses for holders of a driving license or permit]*. https://www.boe.es/eli/es/o/2005/07/28/int2596

Muñiz, J. (2010). Las teorías de los tests: Teoría clásica y teoría de respuesta a los ítems [Test theories: Classical theory and item response theory]. *Papeles del Psicólogo, 31*(1), 57-66.

Navon-Eyal, M., & Taubman-Ben-Ari, O. (2019). Driven by emotions: The association between emotion regulation, forgivingness, and driving styles. *Transportation Research Part F: Traffic Psychology and Behaviour, 65*, 1-9. https://doi.org/10.1016/j.trf.2019.07.005

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, 58*(4), 361-383. https://doi.org/10.1111/jcpp.12675

Pérez-Sánchez, J., Delgado, A. R., & Prieto, G. (2020a). Análisis de los instrumentos empleados en la investigación empírica de la regulación emocional [Analysis of the instruments used in emotion regulation empirical research]. *Cuadernos de Neuropsicología, 14*(1), 165-174. https://doi.org/10.7714/CNPS/14.1.216

Pérez-Sánchez, J., Delgado, A. R., & Prieto, G. (2020b). Propiedades psicométricas de las puntuaciones de los test más empleados en la evaluación de la regulación emocional [Psychometric Properties of the Scores of the Most Common Used Tests in the Evaluation of Emotion Regulation]. *Papeles del Psicólogo, 41*(2), 116-124. https://doi.org/10.23923/pap.psicol2020.2931

Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch [Rasch-modelling a test]. *Psicothema, 15*(1), 94-100.

Prieto, G., & Nieto, E. (2014). Influence of DIF on Differences in Performance of Italian and Asian Individuals on a Reading Comprehension Test of Spanish as a Foreign Language. *Journal of Applied Measurement, 15*(2), 176-188.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment, 31*(4), 557-566. https://doi.org/10.1037/pas0000648

Taubman-Ben-Ari, O. (2012). The effects of positive emotion priming on self-reported reckless driving. *Accident Analysis and Prevention, 45*, 718-725. https://doi.org/10.1016/j.aap.2011.09.039

Trógolo, M. A., Melchior, F., & Medrano, L. A. (2014). The role of difficulties in emotion regulation on driving behavior. *Journal of Behavior, Health & Social Issues, 6*(1), 107-117. https://www.scielo.org.mx/pdf/jbhsi/v6n1/v6n1a8.pdf

Zhang, G., Yau, K. K., & Chen, G. (2013). Risk factors associated with traffic violations and accident severity in China. *Accident Analysis and Prevention, 59*, 18-25. https://doi.org/10.1016/j.aap.2013.05.004