6.2.3 Artículo III

García-Casal, J.A., Martínez-Abad, F., Cid-Bartolomé, T., Smith., S.J., Llano-Ordóñez, K., Perea-Bartolomé, M.V., Goñi-Imizcoz, M., Soto-Pérez, F., Franco-Martín, M. (2017). Usability study and pilot validation of a computer-based emotion recognition test for older adults with Alzheimer's disease and amnestic mild cognitive impairment.

Título: Estudio de usabilidad y Validación piloto de un test del reconocimiento de emociones por ordenador para adultos mayores con enfermedad de Alzheimer y deterioro cognitivo leve de tipo amnésico.

Resumen

Introducción:

Existe una disminución de la capacidad de reconocer la expresión facial de las emociones asociada al envejecimiento normal. Algunos estudios indican que existe una pérdida más pronunciada de esta capacidad en personas con enfermedad de Alzheimer (EA) y deterioro cognitivo leve de tipo amnésico (DCLa). Es clínicamente relevante valorar la capacidad de reconocer emociones dada su importancia para el funcionamiento social cotidiano, ya que afecta a las interacciones sociales, y su potencial para contribuir a la detección temprana del deterioro

cognitivo. A pesar de ello, no se ha validado ninguna escala de evaluación de la capacidad de reconocer emociones para esta población.

Objetivos:

Desarrollar el estudio de usabilidad y la validación piloto de una escala de reconocimiento de emociones por ordenador con interfaz de pantalla táctil (Gradior-Afectos) para personas mayores de 64 años sanas, con EA y con DCLa.

Método:

El test fue administrado a 212 participantes (76.37 ± 6.20 años) clasificados en tres grupos (personas sanas, n = 69; EA, n = 84; y DCLa, n = 59) basados en información clínica y una valoración neurológica y neuropsicológica. El test Gradior-afectos original constaba de 91 ítems que evaluaban las 6 emociones básicas (alegría, tristeza, enfado, asco, sorpresa y miedo) y la expresión neutra. La validación piloto incluyó: (1) un estudio de usabilidad para personas mayores con EA, DCLa y sin deterioro cognitivo; (2) un análisis factorial exploratorio; (3) un análisis de consistencia interna y fiabilidad test-retest; y (4) un estudio de validez discriminante, comparando el desempeño de los tres grupos de validación mediante un análisis de las curvas COR (características operativas de los receptores). La usabilidad se valoró a través de un cuestionario para el paciente y el clínico, observación e indicadores cuantitativos.

Resultados:

Los participantes valoraron el test Gradior-Afectos como accesible y fácil de usar. El instrumento presentó una alta consistencia interna a nivel global (*α de Cronbach ordinal* = 0.96). Las correlaciones test-retest fueron significativas y robustas ($r = 0.840$, $p < 0.001$). El análisis factorial exploratorio, forzando la extracción de 7 factores empíricos, apoyó el modelo teórico, ya que las saturaciones más importantes de los ítems de las 6 emociones y la expresión neutra se

localizaron de manera sistemática en factores coincidentes con cada una de las dimensiones evaluadas. El test validado incluyó 53 ítems, más 7 ítems iniciales de práctica (uno por emoción y uno para la expresión neutra). La versión validada de 53 ítems presentó una alta correlación con la versión extensa de 91 ítems (r = 0.964; p < 0.001). El análisis de las curvas COR indicó que Gradior-Afectos fue capaz de discriminar población sana de personas con EA y DCLa, pero no personas con EA de personas con DCLa. El análisis de regresión ordinal indicó que la puntuación total de Gradior-Afectos mejoró el poder predictivo del MMSE para detectar el grupo diagnóstico de 0.547 a 0.560 (Cox & Snell R^2, p = 0.012), y la velocidad de procesamiento de Gradior-Afectos lo mejoró de 0.547 a 0.563 (Cox & Snell R^2, p = 0.010).

Conclusiones:

Gradior-Afectos es un instrumento válido para la evaluación de la capacidad de reconocer las emociones faciales en personas mayores de 65 años con y sin deterioro cognitivo, e incluso puede resultar de interés para discriminar población sana de personas con algún tipo de deterioro cognitivo. Se recomienda la inclusión de la evaluación de la capacidad de reconocer emociones dentro de las baterías de cribado para la EA y el DCLa, dado que podría mejorar su sensibilidad y especificidad diagnóstica.

Palabras clave: emociones, afectos, demencia, enfermedad de Alzheimer, disfunción cognitiva, neuropsicología, evaluación geriátrica.

Usability study and pilot validation of a computer-based emotion recognition test for older adults with Alzheimer's disease and amnestic mild cognitive impairment

DClinPsy J. Antonio García-Casal[1,2], Dr. Fernando Martínez-Abad[3], Ing. Teresa Cid-Bartolomé[4], Dr. Sarah Jane Smith[5], DClinPsy Katia Llano-Ordóñez[6], Prof. M. Victoria Perea-Bartolomé[1], Md. Miguel Goñi-Imizcoz[7], Dr. Felipe Soto-Pérez[2], Prof. Manuel Franco-Martín[8]

[1] Department of basic psychology, psychobiology and behavioural sciences methods. University of Salamanca, Salamanca, 37005, Spain.

[2] Research & Development Department, Iberian Research Psycho-sciences Institute, INTRAS Foundation, Zamora, 49024, Spain.

[3] Institute of Education Sciences. University of Salamanca, Salamanca, 37008, Spain.

[4] Development Department, INTRAS Foundation, Valladolid, 47016, Spain.

[5] School of Dementia Studies, University of Bradford, Bradford, BD7 1DP, United Kingdom.

[6]Department of Psychiatry, Burgos University Hospital, Burgos, 09006, Spain.

[7]Department of Neurology, Burgos University Hospital, Burgos, 09006, Spain.

[8] Department of Psychiatry and Mental Health, Zamora Regional Hospital, Zamora, 49021, Spain.

Authorship

JAGC and MFM developed the original idea. FMA carried out the statistical analysis. TCB and FSP supervised the technological aspects. KLLO and MGI coordinated the

trial and collected the field data. SJS, MAFM and MVPB contributed to the drafts of the paper and provided valuable comments during the process of writing this manuscript.

Disclosure of interests and acknowledgements

The authors have no competing interests. This work has been supported by the Iberian Research Institute of Psycho-sciences (IBIP, INTRAS Foundation) and the University of Salamanca. We thank all participants and carers for accepting to take part in the study, and we would like to thank Nurse Visitación de las Heras, from the Department of Neurology of Burgos University Hospital, and Neuropsychologist María José Viñas, from INTRAS Foundation´ Memory Clinic, who played a key role in the recruitment of the participants. JAGC, the corresponding author, had full access to all the data in the study and has final responsibility in the decision to submit it for publication.

Corresponding author at:

J. Antonio García-Casal

Carretera de la Hiniesta 137. 49024, Zamora. Spain

Phone number: 0034 608 768 553 / Fax number: 0034 980 557 104

E-mail: agarcia@cop.es

Abstract

Objectives: The aim of this study was to carry out the pilot validation of Affect-Gradior, a computer-based emotion recognition test, with older adults and to evaluate its usability, reliability and validity for the screening of people with Alzheimer´s disease (AD) and amnestic mild cognitive impairment (aMCI).

Methods: The test was administered to 212 participants (76.37 ± 6.20 years) classified into three groups (healthy controls, n = 69; AD, n = 84; and aMCI, n = 59) based on clinical data and formal neurological and neuropsychological assessments. Data on usability were collected by means of a questionnaire and automated evaluation.

Results: The validated test comprised 53 stimuli and 7 practice items (one per emotion). Participants valued Affect-Gradior as accessible and user-friendly. It had high internal consistency (ordinal Cronbach's $\alpha = 0.96$). Test-retest reliability correlations were significant and robust ($r = 0.840$, $p < 0.001$). Exploratory factor analysis supported a seven-factor model of the emotions assessed (neutral expression, happiness, surprise, disgust, sadness, anger and fear). Receiver operating characteristic curve analyses suggested that the test discriminated healthy older adults from AD and aMCI cases. Correct answer score improved MMSE predictive power from 0.547 to 0.560 (Cox & Snell R^2, $p = 0.012$), and Affect-Gradior speed of processing score improved MMSE predictive power from 0.547 to 0.563 (Cox & Snell R^2, $p = 0.010$).

Conclusions: Affect-Gradior is a valid instrument for the assessment of the facial recognition of emotions in older adults with and without cognitive impairment.

Keywords: emotion, affect, Alzheimer disease, mild cognitive impairment, neuropsychology.

Usability study and pilot validation of a computer-based emotion recognition test for older adults with Alzheimer's disease and amnestic mild cognitive impairment.

Introduction

An emotion can be conceptualized as a concerted, generally adaptive, phasic change in multiple physiological systems (including both somatic and neural components) in response to the value of a stimulus (Adolphs, 2002). The capacity to decode facial emotional expressions is probably one of the most powerful vectors of nonverbal communication (Bediou et al., 2009) and an impairment in this ability may lead to social dysfunction and difficulties in interpersonal comprehension (Shimokawa et al., 2001).

The ability of people with mild Alzheimer´s disease (AD) to recognize facial emotional expression can be improved through specific rehabilitation programs (García-Casal et al., 2017). Therefore, the timely detection of emotion recognition deficits could improve the access to treatments having an impact in the quality of life of those affected. Since emotion recognition abilities can also help identify people with amnestic mild cognitive impairment (aMCI) and AD (Bertoux et al., 2015; Bora, Velakoulis, & Walterfang, 2016), emotion recognition assessment could be useful both for early screening and for differential diagnosis. The early detection of dementia is a key aspect to initiate timely treatments and to reduce morbidity (Huntley, Gould, Liu, Smith, & Howard, 2015; Laske et al., 2015).

A loss in capacity for facial emotion recognition is associated with normal ageing, especially for negative emotions like anger, sadness and fear (Orgeta, 2010). A more pronounced deficit in ability to identify emotional facial expressions has also been

found in people with AD (Kumfor et al., 2014; Sapey-Triomphe et al., 2015; Tabernero, Rubinstein, Cossini, & Politis, 2016) and aMCI (McCade, Savage, & Naismith, 2011), and has been related to the evolution of aMCI into AD (Bediou, et al., 2009). Processing speed is one of the cognitive functions that undergo the greatest declines in ageing and is specifically impaired in AD (Garcia-Rodriguez, Ellgring, Fusari, & Frank, 2009) and aMCI (Varjassyová et al., 2013), however emotion recognition processing speed has not been investigated.

The study of emotions has found six basic emotions that are universally identified: happiness, disgust, fear, surprise, sadness and anger (Ekman and Friesen, 1971). Emotion recognition capacity assessment based on pictures has been studied with three types of tasks (McLellan, Johnston, Dalrymple-Alford, & Porter, 2008): identification of facial expressions (the participant is required to choose which emotion label describes the expression shown in a photo), discrimination of facial expressions (the participant looks at pairs of photos and indicates if the expressions are the same or different), and matching and selecting facial expressions (the participant matches a target expression to one of several alternatives or selects a target expression). Ekman and Friesen, developed a system to encode visibly different facial movements: "The Facial Action Code" (Ekman and Friesen, 1976) and designed a classic emotion recognition assessment test that required participants to choose which of the six emotion labels best described an emotion. This test, The Ekman Faces, is the most commonly used. However, it is subject to limitations; namely a) it only provides outcomes about correct and incorrect answers, not about processing speed and the nature of errors; and b) it has not been validated for older adults. Consulting with people with cognitive impairment is crucial to ensure usability in the design of technological solutions (Span, Hettinga, Vernooij-Dassen, Eefsting, & Smits, 2013). There is a lack of outcome

measures specifically validated for older adults with cognitive impairment which poses serious concerns about their suitability for this population (Meiland et al., 2017), with the consequent risk of misdiagnosis (Shenoy and Harugeri, 2015). Further tests developed for this purpose have similar limitations, e.g. the Penn Emotion Recognition Test (Gur et al., 2002), Izard photographs (Allender and Kaszniak, 1989), the Florida Affect Battery (Cadieux and Greve, 1997), and the International Affect Picture System (Lang, Bradley, & Cuthbert, 1999).

The current study is designed to address the limitations of extant emotional recognition tests by drawing on opportunities brought about by advances in technology. Affect-Gradior, a picture based computerized test of emotion recognition, was studied for its (1) usability in older adults with AD and aMCI and without cognitive impairment; (2) exploratory factorial analysis; (3) reliability, assessing its internal consistency and test-retest stability; and (4) discriminant validity, comparing the performance of people with AD and aMCI to that of healthy controls.

Methods

The study was a cross sectional descriptive study. We adhered to STARD and STROBE guidelines; the supporting checklists are available as supplemental online material 1.

Participants

Affect-Gradior was administered to 212 participants (M = 76.37; ± 6.20 years): 69 healthy older adults (M = 73.14 ± 6.28 years); 84 people with AD (M = 78.27 ± 5.81 years); and 59 people with aMCI (M = 77.60 ± 5.01 years). The diagnoses of dementia and aMCI were provided by a clinical psychologist and two neurologists blind to the objectives of the study based on detailed neurological, neuropsychological, laboratory and neuro-imaging (Structural Magnetic Resonance Image) data from each participant.

Diagnosis of dementia was determined using the DSM-IV-TR criteria (APA, 2000); if these criteria were met, the neurologist determined the specific type of dementia using the revised NINCDS-ADRDA criteria for AD (Dubois et al., 2007). The diagnosis of aMCI was established according to the consensus criteria of the International Working Group on aMCI (Winblad et al., 2004). Participants with comorbid psychiatric conditions were excluded from the study. The healthy older adults were recruited from accompanying persons with no blood relationship with the patient and without pre-existing neurological disorders that could potentially cause neuropsychological deficits (e.g. stroke, epilepsy, movement disorder, brain tumour or severe head trauma). All the healthy controls (HC) were judged to be cognitively normal by the clinical psychologist and the neurologist.

Procedures

The study participants were consecutively recruited based on selection criteria from the outpatients' memory clinics at Burgos University Hospital in Burgos and INTRAS Foundation in Zamora, following protocol approval by the local ethical committee (CEIC Ref.1381) between October 2014 and September 2016. All of the participants were blind to the objectives of the assessment, and gave their informed consent prior to inclusion in the study. Assessments were carried out by four clinical psychologists and three neurologists who were blind to the objectives of the study.

Usability was assessed through an iterative process. With the findings from each cycle, the test instructions, length of the test and practice trial were revised. A pilot version of Affect-Gradior was applied to 6 people with AD and 6 people with aMCI to gather suggestions regarding accessibility and usability. The suggestions formulated by the participants were: to add written instructions, to add a training item, to rename some of

the emotions, to remove the music from the instructions as it made it difficult to understand them and to keep the same order in the emotions´ tags. With these findings the test instructions, practice trials and tags were revised to make them more accessible. Participants had a proxy (primary caregiver) who was interviewed prior to neuropsychological assessment. All the participants completed Affect-Gradior and MMSE. Most of those with AD and aMCI also underwent a complete neuropsychological assessment that lasted approximately one hour. Test-retest reliability was examined using data from participants who received Affect-Gradior test in a consecutive visit (n = 29).

Whilst completing the task participants were asked to position their hands over the table, in front of the touchscreen. The distance between the touchscreen and the face of the person was of 65 cm following the International Organization for Standardization ISO-9241 norms (Woo, White, & Lai, 2016). The clinician could pause the task if needed by pressing a green button at the upper left corner of the screen.

Outcome measures

Affect-Gradior is a touchscreen emotion recognition test available on Spanish and English languages. The test required participants to identify the correct emotion from six basic emotions and a neutral expression consisting of 91 stimuli, 13 per emotion (Figure 1). The emotional stimuli comprised colour photographs of professional actors expressing six basic emotions and a neutral expression. The expressions were depicted by 13 different actors (6 male and 7 female) photographed by a professional photographer with a 10 megapixel digital camera. The test was designed using Microsoft Visual Studio V6 software and Visual Basic programming language.

The instructions were presented displayed on a 13´3 inch screen and could be heard through loudspeakers. To eliminate the factors related to oral verbal processing, the participants were required to match the images of stimuli and a written label, avoiding the use of spoken language. The task of participants was to respond via the touchscreen pressing the label that best described the facial expression shown. The pictures measured 7 x 11 cm, the labels displaying the emotions measured 7 x 2 cm (See Figure 1). If the participant did not provide an answer after 33 seconds the test automatically moved to the next stimuli, and recorded an error of omission. A trial run consisting of a practice item preceded the test. The program allowed the sociodemographic and clinical data linked to each person to be recorded. All the participant data was exported to an Access or Excel file comprising the following information:

- Total scores and partial scores per emotion reflecting the correctly identified emotions.
- Errors of commission (the participant chose the wrong answer) and errors of omission (the participant did not provide any answer).
- Total emotion processing speed and processing speed per emotion (([correct answers-(omission + commission errors)]/(reaction time/1000)), with a correction for negative values: (([correct answers-(omission + commission errors)]*(reaction time/1000))).
- Total precision of processing ([correct answers-commission errors]/91)*100).
- Type of answer provided instead of the correct answer (e.g. the person chose sadness instead of anger).
- Identity and gender of the poser depicting each stimuli.

General cognitive capacity was assessed with the Spanish version of the Mini-mental State Examination (MMSE) (Lobo et al., 1999). A more comprehensive neuropsychological assessment of cognitive status was carried out with CAMCOG-R (Roth, Huppert, Mountjoy, & Tym, 1998). Mood was assessed with the short version of the Geriatric Depression Scale (GDS-D) (Martín et al., 2002; Sheikh and Yesavage, 1986). Usability was assessed with a questionnaire available as supplemental online material 2. Two objective markers of usability were used: the trend line of the percentage of correct answers along the test and the distribution of the discriminative power of the items within the test.

Statistical analysis

Baseline characteristics of the three groups were compared using independent F, t and X^2 tests as appropriate, depending on the distribution of the variables. An analysis of variance (ANOVA) was conducted to compare results of Affect-Gradior by diagnosis category. Post hoc analysis where carried out with Bonferroni correction for multiple comparisons. Correlation indexes were used to assess association among all the variables.

Internal consistency was calculated using ordinal Cronbach's α, which reflects the average inter-item correlation score and, as such, will increase when correlations between the items increase (Bland and Altman, 1997). Test-retest reliability comparisons were conducted using Pearson's correlation as stability coefficient to assess consistency of the data collected. Exploratory Factor Analysis was used to arrange the variables in domains using principal components extraction. The factorial analysis included the extraction of seven factors, and the result was rotated under Varimax procedure to ease interpretation.

A receiver-operating characteristic (ROC) curve analysis was performed to determine if Affect-Gradior could discriminate between healthy older adults, people with aMCI and people with AD. The optimal cut-off scores of the Affect-Gradior test for the discrimination between patients and HC was determined using Youden's index, considering consensus diagnosis as the gold standard. Among all subjects multiple ordinal regression analyses (Harrell, 2015), enter method, were performed to identify independent predictors for diagnosis and to analyze if Affect-Gradior test contributed variance above and beyond MMSE in the detection of AD and aMCI.

Results

The sociodemographic and clinical data of the participants are summarized in Table 1. The three groups were equivalent in civil status, education and mood. There was a significant difference in age between healthy participants, people with aMCI and people with AD. In a simple regression analysis, age had a predictive power over correct answers ($R^2 = 0.055$; $t = -3.497$; $p = 0.001$), but not over processing speed ($R^2 = 0.013$; $t = -1,694$; $p = 0.092$). To determine if age explained the variance of Affect-Gradior outcomes, a multiple regression analysis (enter method) was performed with total correct answers and total emotion recognition processing speed as dependent variables and the participant based variables as independent variables (age and group). The resulting regression model excluded age as a significant factor. The diagnostic group (HC, AD or aMCI) explained 21% of the variance of total correct answers ($R^2 = 0.206$; $F = 27.068$; $p < 0.001$), and age had no predictive power ($\beta = -0.124$; $t = -1.376$; $p = 0.170$). In the case of emotion recognition processing speed, the group explained 14% of the variance ($R^2 = 0.141$; $F = 17.143$; $p < 0.001$) and age had no predictive power ($\beta = 0.033$; $t = 0.240$; $p = 0.811$). As a consequence, age was not entered as a covariate in the analysis.

The Affect-Gradior total score was significantly correlated with the MMSE ($r= 0.487$, $p < 0.001$) and the CAMCOG-R ($r= 0.432$, $p < 0.001$), showing better emotion recognition at higher cognitive scores (Table 2). The Affect-Gradior total emotion processing speed score was significantly correlated with the MMSE ($r = 0.447$; $p < 0.001$), but showed no significative correlations with the CAMCOG-R ($r = 0.251$; $p = 0.051$). Correlations for individual emotions and processing speeds varied. The Affect-Gradior scores showed no significant correlations with GDS-D scale neither for emotion recognition ($r = -0.091$; $p = 0.314$) nor for emotions processing speed ($r = 0.013$; $p = 0.881$).

Content Validity and Usability

Usability was assessed with a sample of 27 participants ($M = 76.81 \pm 5.65$ years): 11 healthy older adults ($M = 76.73 \pm 5.57$ years); 8 people with AD ($M = 79.25 \pm 6.32$ years); and 8 people with aMCI ($M = 74.50 \pm 4.63$ years). There were no significant differences between this sample and the rest of the participants in age and education ($p > 0.05$).

The qualitative data were categorized in clusters for a semantic analysis of frequencies. All the participants found the test easy to use, amusing or entertaining, and two of them (7%) spontaneously expressed their satisfaction for having used a computer for the first time. Regarding the duration of the test, 23 participants (85%) found it appropriate; one found it repetitive and 3 (11%) found it too long. When asked about the clarity of the instructions, all the participants said that they were clear and 4 (15%) suggested that they should warn that the actors enact different emotions, therefore appearing more than one time. Thirteen participants (48%) answered that they had been able to identify all the emotions. All the participants found working with the touchscreen easy, and 3 of

them (11%) added that after this experience they felt encouraged to continue using it. The researchers detected that in 3 cases (11%) the participants needed assistance at the beginning of the test as they did not release the touchscreen after pressing it, and 1 participant should be asked not to grab the edge of the touchscreen so that the answers could be registered.

The quantitative data of the usability questionnaire are summarized in Table 3. The three groups of participants considered the test easy and comprehensible and had little previous experience with computers. There were no significant differences in the perception of the difficulty ($Z = -1.27$, $p = 0.21$) and comprehensiveness ($Z = -1.39$, $p = 0.16$) of the test between the participants and the researchers. However, the researchers observed that participants with AD found the test less easy to complete and comprehend than HC and participants with aMCI.

The average time of completion in minutes for the 91 items scale was 15.82 (Standard Deviation (SD) = 6.54) for people with AD, 14.10 (SD = 4.67) for people with aMCI and 10.86 (SD = 2.68) for healthy controls. The trend line of the percentage of correct answers for the 91 items showed a downtrend (Figure 2a), while the trend line for the first 60 items showed an uptrend (Figure 2b). The distribution of the discriminative power of the items within the test was not even: of the 38 items that had low discriminative power, 14 (37%) were within the first 20 items and 13 (34%) were within the last 20 items (Supplemental online material 3).

Reliability

The internal consistency value for the whole instrument was high (Cronbach's $\alpha = 0.87$; ordinal Cronbach's $\alpha = 0.96$). The mean time between the first and second assessment for the test-retest reliability study was of 150.26 days (SD = 70.36). The intra-class

correlation coefficient between the scores at baseline and retest was high ($r = 0.840$; $p < 0.001$). The mean difference between the pre-test and post-test scores was not statistically significant ($t = 0.737$; $p = 0.468$).

Factorial validity

Exploratory factor analysis was used to arrange the variables in domains using principal components extraction. The analysis showed that the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.762 indicating that the data were appropriate for this analysis. Bartlett's test of sphericity was significant ($X^2 = 3366.396$ $p < 0.000$), indicating that correlations existed among some of the items. The estimations were based on Pearson's correlations between indicators using a principal component analysis as suggested by Marsh et al. (Marsh, Morin, Parker, & Kaur, 2014) and later replicated with a tetrachoric correlation matrix as suggested by Osborne & Fitzpatrick (Osborne and Fitzpatrick, 2012). The rotation converged in 7 iterations which together accounted for 60% of the variance in the tetrachoric model (Table 4), all the items of each component belonged to the same emotion. To enhance interpretability, only factor loadings ≥ 0.3 were selected, this process left 53 items with enough discriminative power. Four items had a second factor load, but the saturation was never higher than the one that belonged to their main factor. Supplemental online material 4 shows the seven-component rotated solution with the factor loadings of each item. The final model comprised 53 items and 7 factors that fitted the six basic emotions and the neutral expression: neutral expression (11 items), happiness (8 items), surprise (9 items), disgust (7 items), sadness (8 items), anger (5 items) and fear (5 items). All the calculations of the psychometric properties of the test were performed over the final 53 items version obtained after the exploratory factor analysis. The 53 items version had a high correlation with the 91 items version ($r = 0.964$; $p < 0.001$).

Discriminant validity

Table 1 summarizes the total emotion recognition and processing speed results for each group, as well as the partial scores for each emotion. Total emotion recognition scores differed significantly between the three groups for total correct answers ($F_{2,209} = 26.244$; $p < 0.001$), happiness ($F_{2,209} = 4.635$; $p = 0.011$), disgust ($F_{2,209} = 8.506$; $p < 0.001$), anger ($F_{2,209} = 6.239$; $p = 0.002$), neutral ($F_{2,209} = 17.250$; $p < 0.001$), surprise ($F_{2,209} = 22.354$; $p < 0.001$), errors of omission ($F_{2,209} = 4.708$; $p = 0.01$), errors of commission ($F_{2,209} = 19.135$; $p < 0.001$), mean reaction time ($F_{2,209} = 20.206$; $p < 0.001$), precision of processing ($F_{2,209} = 26.244$; $p < 0.001$) and processing speed ($F_{2,209} = 17.153$; $p < 0.001$). Post hoc analysis with the Bonferroni test adjusted for multiple comparisons showed significant differences between the three groups in total correct answers, precision of processing, and processing speed which were the three variables with the highest discriminative power.

To evaluate the screening accuracy of the Affect-Gradior test, ROC curve analyses were performed to compare pairs of diagnostic groups (HC×AD, ADxaMCI and HC×aMCI) with the total emotion recognition and processing speed scores, selecting the optimal cut-off scores. Precision of processing was excluded from this analysis because it is a variable derivative of correct answers, with the same discriminative power. The results of the ROC curve analysis of the Affect-Gradior are displayed in Figure 3 and Table 5.

Total correct answers discrimination between HC and AD (AUC = 0.791) was better than that between HC and aMCI (AUC = 0.717) and between aMCI and AD (AUC = 0.642). Processing speed discrimination between HC and AD (AUC = 0.829) was better than the one between HC and aMCI (AUC = 0.748) and between aMCI and AD (AUC = 0.670). For all groups emotion recognition processing speed gave better

discrimination than correct answers. The optimal balances between sensitivity and specificity for the three variables and groups, as well as AUC 95% confidence intervals are displayed in Table 5.

The multiple ordinal regression analysis (enter method) showed that Affect-Gradior correct answer score improved MMSE predictive power from 0.547 to 0.560 (Cox & Snell R^2, p = 0.012), and Affect-Gradior speed of processing score improved MMSE predictive power from 0.547 to 0.563 (Cox & Snell R^2, p = 0.010).

Discussion

The aim of this study was to determine the psychometric properties and usability of Affect-Gradior as an emotion recognition test, and to assess its accuracy in the screening of AD and aMCI. To our knowledge, this is the first study to validate an emotion recognition test in a sample of older adults. The present study suggests that the battery has good psychometric characteristics and high usability and acceptability among older adults with and without cognitive impairment.

Psychometric properties

The Affect-Gradior test presented good internal consistency and test-retest intraclass correlations, thereby suggesting that it is a reliable scale for evaluating emotion recognition in older adults. The results of the exploratory factor analysis supported the hypothesis that the factorial structure reflects the seven emotions assessed, as all the items included in each factor belonged to the same emotion. In the final model each emotion had a different load, depending on the discriminative power detected in the exploratory factor analysis.

The total correct answers and processing speed scores were statistically different between people with AD, aMCI and healthy older adults, suggesting that there are detectable differences between groups in specific emotion recognition abilities. Affect-Gradior correct answers and processing speed scores had a good discrimination between HC and AD, and between HC and aMCI, but a poor discrimination between aMCI and AD. The multiple regression analysis showed that Affect-Gradior improved the diagnostic accuracy of MMSE, while simultaneously providing a measure of emotion recognition. The processing speed associated to emotion recognition seems to share the slowing of processing speed that occurs very early in the path leading to dementia (Welmer, Rizzuto, Qiu, Caracciolo, & Laukka, 2014).

Assessments of ability in emotion recognition often rely on clinical judgment with a high risk of proxy bias. The availability of a cut-off score, which has a known sensitivity and specificity, to discriminate between patients with AD, aMCI and healthy older adults could allow for a more precise definition of impairment in emotion recognition in people with cognitive impairment. Emotion recognition assessments could be included in screening for dementia to overcome limitations of cognitive instruments and to enable access for people with emotion recognition deficits to specific emotion recognition rehabilitation interventions. Affect-Gradior could also be a valid pre-post assessment tool for emotion recognition rehabilitation programs.

Usability

All assessed participants considered the test accessible, amusing or easy to use, and liked working with the touchscreen despite the fact that they had little or no experience with computers. In fact, some of them verbalized that after the experience they felt encouraged to continue using computers. Incidences registered by the administrators

(difficulty to release the touchscreen after pressing it and tendency to grab it by its edge) should be considered by professionals using touchscreens with older adults, so that they can provide them with adequate guidance if they need it.

Test instructions were made more accessible including older adults´ recommendations (e.g. warning that some actors appeared more than once). The downtrend of the percentage of the correct answers in the 91 items test was interpreted as a sign of tiredness, as the participants´ performance tended to get worst as the test progressed. As a matter of fact, the trend line for the first 60 items was positive, supporting this hypothesis. In addition, most of the items with low discriminative power concentrated at the beginning and at the end of the test. The high rate of low discriminative items at the beginning might be interpreted as a lack of training effect, suggesting that training items for all the emotions should be included in the beginning, for that reason, the final version of the test included 7 training items, 1 per emotion, selected based on their high predictive power. On the other hand, the high rate of low discriminative items at the end of the test might be interpreted as a sign of tiredness of the participants, in line with the correct answers´ trend line. This information supports the decision of building a shorter version of the test, improving its usability. The necessity of including tests of performance validity in the batteries has been highlighted, as the validity of the assessment relies on the examinee's full motivation and effort to perform as well as possible (Roebuck-Spencer, Vincent, Gilliland, Johnson, & Cooper, 2013). It has also been suggested that automated evaluation mechanisms should be adopted to improve the empirical methods employed to assess usability (Baez et al., 2013). We recommend the inclusion of the distribution of the discriminative power of the items and the trend line of the percentage of correct answers as automated and objective performance validity and usability tests.

Facial emotion recognition in AD and aMCI

Our results indicate that facial emotion recognition is poorer in people with AD and aMCI, compared with healthy older adults. These results correspond with previous studies reporting deficits in facial emotion recognition in people with AD (Henry et al., 2012; Kumfor, et al., 2014; Sapey-Triomphe, et al., 2015; Tabernero, et al., 2016) and people with aMCI (McCade et al., 2013; Pietschnig et al., 2016; Varjassyová, et al., 2013). Conversely, other studies found a preserved emotion recognition capacity in people with AD (Burnham and Hogervorst, 2004; Freedman, Binns, Black, Murphy, & Stuss, 2013; Hsieh, Hornberger, Piguet, & Hodges, 2012). The wealth of data available about emotion recognition in AD and MCI, to which this study adds, justifies carrying out a meta-analysis to derive a consensus with regards to this data; as carried out previously in the field of frontotemporal dementia (Bora, et al., 2016) and healthy older adults (Ruffman, Henry, Livingstone, & Phillips, 2008).

Deficits were progressively evident in aMCI to AD, starting with a deficit in disgust and surprise in aMCI to a deficit in disgust, happiness, anger, surprise and the neutral expression in AD. This could be an effect of the progressive degeneration of brain structures modulating facial emotion recognition, this hypothesis is coherent with previous findings (Spoletini et al., 2008). All participants had more difficulties in recognizing negative emotions than happiness and surprise, this results confirm previous studies where older adults have been reported to have a deteriorated capacity to recognize negative emotions (Orgeta, 2010; Sarabia-Cobo, Navas, Ellgring, & García-Rodríguez, 2015; Tabernero, et al., 2016). The results for single emotions mirror this distinction between positive and negative emotions, with higher recognition rates for positive affective states (happiness = 90% and surprise = 63%) and lower

recognition rates for negative emotions, of which anger (39%) and fear (20%) represented the opposite extremes.

Our results are consistent with previous investigations that found a correlation between emotion recognition and cognitive status. Some of those studies suggested that emotion recognition capacity is mediated by cognitive status (Bertoux, et al., 2015; Miller et al., 2012; Torres et al., 2015). However, poor performance on the Affect-Gradior cannot reliably be explained by crude cognitive deficits captured by the MMSE or CAMCOG-R. The relationship between emotion recognition and cognitive deterioration is yet to be established, i.e. whether emotion recognition deficits are primary or secondary to cognitive deterioration. Our hypothesis is that both emotion recognition capacity and general cognition are affected by the neurodegeneration and neurophysiological changes associated to AD and aMCI; future research should investigate this hypothesis more fully. Our study contradicts previous findings that found no correlation between recognition of facial expressions of emotion and general cognition measured with the MMSE (Bediou, et al., 2009; Shimokawa, et al., 2001).

Affect-Gradior showed no significant correlations with GDS-D mood test, suggesting that emotion recognition is independent of mood in non-depressed older adults, confirming previous findings (Orgeta, 2014; Torres, et al., 2015).

Conclusions

Affect-Gradior is a valid instrument for the assessment of the facial recognition of emotions in older adults with and without cognitive impairment and has adequate psychometric characteristics. This study proposes the inclusion of emotion recognition tests in screening for dementia and aMCI based on improved diagnostic accuracy. In conclusion, the Affect-Gradior test may prove useful for both clinical and research

purposes to investigate global emotion recognition ability as well as selective impairment of individual basic emotions recognition in older adults. A confirmatory factorial analysis and a final validation with a normative study of the test should be carried out with a wider sample.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews, 1*(1), pp. 21-62.

Allender, J., & Kaszniak, A. W. (1989). Processing of emotional cues in patients with dementia of the Alzheimer's type. *International Journal of Neuroscience, 46*(3-4), pp. 147-155.

APA, A. P. A. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR fourth edition (text revision)*. Washington: American Psychological Association.

Baez, S., Couto, B., Herrera, E., Bocanegra, Y., Trujillo-Orrego, N., Madrigal-Zapata, L., . . . Villegas, A. (2013). Tracking the cognitive, social, and neuroanatomical profile in early neurodegeneration: Type III Cockayne syndrome. *Frontiers in Aging Neuroscience, 5,* p 80.

Bediou, B., Ryff, I., Mercier, B., Milliery, M., He, M.-A., d'Amato, T., . . . Krolak-Salmon, P. (2009). Impaired social cognition in mild Alzheimer disease. *Journal of Geriatric Psychiatry and Neurology*, *22*(2), pp. 130-140.

Bertoux, M., de Souza, L. C., Sarazin, M., Funkiewiez, A., Dubois, B., & Hornberger, M. (2015). How preserved is emotion recognition in Alzheimer disease compared with behavioral variant frontotemporal dementia? *Alzheimer Disease & Associated Disorders, 29*(2), pp. 154-157.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj, 314*(7080), p 572.

Bora, E., Velakoulis, D., & Walterfang, M. (2016). Meta-Analysis of Facial Emotion Recognition in Behavioral Variant Frontotemporal Dementia Comparison With

Alzheimer Disease and Healthy Controls. *Journal of Geriatric Psychiatry and Neurology*, *29(4), pp. 205-211.*

Burnham, H., & Hogervorst, E. (2004). Recognition of Facial Expressions of Emotion by Patients with Dementia of the Alzheimer Type. *Dementia and Geriatric Cognitive Disorders, 18*(1), pp. 75-79.

Cadieux, N. L., & Greve, K. W. (1997). Emotion processing in Alzheimer's disease. *Journal of the International Neuropsychological Society, 3*(05), pp. 411-419.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., . . . Jicha, G. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology, 6*(8), pp. 734-746.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), p 124.

Ekman, P., & Friesen, W. (1976). Measuring facial movement. [journal article]. *Environmental psychology and nonverbal behavior, 1*(1), pp. 56-75.

Freedman, M., Binns, M. A., Black, S. E., Murphy, C., & Stuss, D. T. (2013). Theory of mind and recognition of facial emotion in dementia: Challenge to current concepts. *Alzheimer Disease and Associated Disorders, 27*(1), pp. 56-61.

García-Casal, J. A., Goñi-Imizcoz, M., Perea-Bartolomé, M., Soto-Pérez, F., Smith, S. J., Calvo-Simal, S., & Franco-Martín, M. (2017). The Efficacy of Emotion Recognition Rehabilitation for People with Alzheimer’s Disease. *Journal of Alzheimer's Disease, 57*(3), pp. 937-951.

Garcia-Rodriguez, B., Ellgring, H., Fusari, A., & Frank, A. (2009). The role of interference in identification of emotional facial expressions in normal ageing and dementia. *European Journal of Cognitive Psychology, 21*(2-3), pp. 428-444.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., . . . Gur, R. E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods, 115*(2), pp. 137-143.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis* Nashville: Springer.

Henry, J. D., Thompson, C., Rendell, P. G., Phillips, L. H., Carbert, J., Sachdev, P., & Brodaty, H. (2012). Perception of biological motion and emotion in mild cognitive impairment and dementia. *Journal of the International Neuropsychological Society, 18*(5), pp. 866-873.

Hsieh, S., Hornberger, M., Piguet, O., & Hodges, J. R. (2012). Brain correlates of musical and facial emotion recognition: Evidence from the dementias. *Neuropsychologia, 50*(8), pp. 1814-1822.

Huntley, J., Gould, R., Liu, K., Smith, M., & Howard, R. (2015). Do cognitive interventions improve general cognition in dementia? A meta-analysis and meta-regression. *Bmj Open, 5*(4), p e005247.

Kumfor, F., Sapey-Triomphe, L.-A., Leyton, C. E., Burrell, J. R., Hodges, J. R., & Piguet, O. (2014). Degradation of emotion processing ability in corticobasal syndrome and Alzheimer's disease. *Brain, 137*(11), pp. 3061-3072.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*

Laske, C., Sohrabi, H. R., Frost, S. M., Lopez-de-Ipina, K., Garrard, P., Buscema, M., . . . O'Bryant, S. E. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimers Dement, 11*(5), pp. 561-578.

Lobo, A., Saz, P., Marcos, G., Día, J., De La Cámara, C., Ventura, T., . . . Aznar, S. (1999). Revalidation and standardization of the cognition mini-exam (first Spanish version of the Mini-Mental Status Examination) in the general geriatric population. *Medicina Clinica, 112*(20), pp. 767-774.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, pp. 85-110.

Martín, F. S., Andrade, C., Molina, J., Muñoz, P., Carretero, B., Rodríguez, M., & Silva, A. (2002). Validation of the Spanish version of the geriatric depression scale (GDS) in primary care. *International Journal of Geriatric Psychiatry, 17*(3), pp. 279-287.

McCade, D., Savage, G., Guastella, A., Hickie, I. B., Lewis, S. J. G., & Naismith, S. L. (2013). Emotion recognition in mild cognitive impairment: Relationship to psychosocial disability and caregiver burden. *Journal of Geriatric Psychiatry and Neurology, 26*(3), pp. 165-173.

McCade, D., Savage, G., & Naismith, S. L. (2011). Review of emotion recognition in mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders, 32*(4), pp. 257-266.

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2008). The recognition of facial expressions of emotion in Alzheimer's disease: A review of findings. *Acta Neuropsychiatrica, 20*(5), pp. 236-250.

Meiland, F., Innes, A., Mountain, G., Robinson, L., van der Roest, H., García-Casal, J. A., . . . Franco-Martin, M. (2017). Technologies to Support Community-Dwelling Persons With Dementia: A Position Paper on Issues Regarding

Development, Usability, Effectiveness and Cost-Effectiveness, Deployment, and Ethics. [Original Paper]. *JMIR Rehabil Assist Technol, 4*(1), p e1.

Miller, L. A., Hsieh, S., Lah, S., Savage, S., Hodges, J. R., & Piguet, O. (2012). One size does not fit all: face emotion processing impairments in semantic dementia, behavioural-variant frontotemporal dementia and Alzheimer's disease are mediated by distinct cognitive deficits. *Behavioural Neurology, 25*(1), pp. 53-60.

Orgeta, V. (2010). Effects of age and task difficulty on recognition of facial affect. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *65*(3), pp. 323-327.

Orgeta, V. (2014). Emotion recognition deficits in late life: effects of mild depressive affect. *Experimental Aging Research, 40*(1), pp. 1-12.

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation, 17*(15), pp. 1-8.

Pietschnig, J., Aigner-Wöber, R., Reischenböck, N., Kryspin-Exner, I., Moser, D., Klug, S., . . . Lehrner, J. (2016). Facial emotion recognition in patients with subjective cognitive decline and mild cognitive impairment. *International Psychogeriatrics, 28*(03), pp. 477-485.

Roebuck-Spencer, T. M., Vincent, A. S., Gilliland, K., Johnson, D. R., & Cooper, D. B. (2013). Initial clinical validation of an embedded performance validity measure within the automated neuropsychological metrics (ANAM). *Archives of Clinical Neuropsychology*, *28*(7), pp. 700-710.

Roth, M., Huppert, F., Mountjoy, C., & Tym, E. (1998). The Cambridge exam for mental disorders of the elderly-revised. London: Cambridge University Press.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews, 32*(4), pp. 863-881.

Sapey-Triomphe, L.-A., Heckemann, R. A., Boublay, N., Dorey, J.-M., Hénaff, M.-A., Rouch, I., . . . Initiative, A. s. D. N. (2015). Neuroanatomical Correlates of Recognizing Face Expressions in Mild Stages of Alzheimer's Disease. *PLoS ONE, 10*(12), p e0143586.

Sarabia-Cobo, C. M., Navas, M. J., Ellgring, H., & García-Rodríguez, B. (2015). Skilful communication: Emotional facial expressions recognition in very old adults. *International Journal of Nursing Studies, 54*, pp. 104-111.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS). *Clinical Gerontologist, 5*(1-2), pp. 165-173.

Shenoy, P., & Harugeri, A. (2015). Elderly patients' participation in clinical trials. *Perspectives in clinical research, 6*(4), p 184.

Shimokawa, A., Yatomi, N., Anamizu, S., Torii, S., Isono, H., Sugai, Y., & Kohno, M. (2001). Influence of deteriorating ability of emotional comprehension on interpersonal behavior in Alzheimer-type dementia. *Brain and Cognition, 47*(3), pp. 423-433.

Span, M., Hettinga, M., Vernooij-Dassen, M., Eefsting, J., & Smits, C. (2013). Involving people with dementia in the development of supportive IT applications: A systematic review. *Ageing Research Reviews, 12*(2), pp. 535-551.

Spoletini, I., Marra, C., Di Iulio, F., Gianni, W., Sancesario, G., Giubilei, F., . . . Spalletta, G. (2008). Facial emotion recognition deficit in amnestic mild

cognitive impairment and Alzheimer disease. *The American Journal of Geriatric Psychiatry, 16*(5), pp. 389-398.

Tabernero, M. E., Rubinstein, W. Y., Cossini, F. C., & Politis, D. G. (2016). Reconocimiento facial de emociones básicas en demencia frontotemporal variante conductual y en enfermedad de Alzheimer. *Neurología Argentina, 8*(1), pp. 8-16.

Torres, B., Santos, R. L., Sousa, M. F. B. d., Simões Neto, J. P., Nogueira, M. M. L., Belfort, T. T., . . . Dourado, M. C. N. (2015). Facial expression recognition in Alzheimer's disease: a longitudinal study. *Arquivos de neuro-psiquiatria, 73*(5), pp. 383-389.

Varjassyová, A., Hořínek, D., Andel, R., Amlerova, J., Laczó, J., Sheardová, K., . . . Bradáč, O. (2013). Recognition of facial emotional expression in amnestic mild cognitive impairment. *Journal of Alzheimer's disease: JAD, 33*(1), p 273.

Welmer, A.-K., Rizzuto, D., Qiu, C., Caracciolo, B., & Laukka, E. J. (2014). Walking Speed, Processing Speed, and Dementia: A Population-Based Longitudinal Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *69*(12), pp. 1503-1510.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., . . . Almkvist, O. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine, 256*(3), pp. 240-246.

Woo, E., White, P., & Lai, C. (2016). Ergonomics standards and guidelines for computer workstation design and the impact on users' health–a review. *Ergonomics, 59*(3), pp. 464-475.

Table 1. Sociodemographic and clinical variables of the participants

	H (n =69)		aMCI (n =59)		AD (n = 84)			Post-hoc analysis		
	n/*M*	%/SD	n/*M*	%/SD	n/*M*	%/SD	*p-value*	AD-H	AD-aMCI	aMCI-H
Age	73.14	6.29	77.60	5.01	78.27	5.81	<0.001	<0.001	-	0.001
Sex										
Man	45	63%	30	53%	52	62%	-			
Woman	26	37%	27	47%	32	38%				
Civil Status										
Partnered	51	72%	38	67%	56	67%	0.747			
Single/widow	20	28%	19	33%	28	33%				
Education	7.76	3.45	7.09	3.89	6.86	2.83	0.207			
GDS	1.07	0.31	2.11	0.99	3.10	0.53	<0.001	<0.001	<0.001	<0.001
CDR	0.04	0.15	0.60	0.20	1.06	0.43	<0.001	<0.001	<0.001	<0.001
GDS-D	3.82	3.51	2.66	3.15	2.92	2.78	0.184			
MMSE	28.38	1.31	24.02	2.36	21.57	3.57	<0.001	<0.001	0.008	<0.001
CAMCOG-R	88.00	15.43	76.16	8.55	66.45	9.29	0.001	0.017	0.050	-
Affect-GRADIOR										
Assessment duration	4' 11''	1' 30''	6' 8''	2' 48''	7' 10'	3' 43''				
Total Correct Answers (53)	32.68	6.79	27.35	6.82	23.81	8.68	<0.001	<0.001	0.048	0.001
Happiness (8)	7.55	1.27	7.30	1.63	6.80	1.74	<0.001	<0.001	0.007	-
Disgust (7)	3.62	2.08	2.74	1.82	2.39	1.73	0.001	0.001	-	0.047
Anger (5)	2.41	1.39	1.84	1.42	1.63	1.37	0.003	0.002	-	-
Fear (5)	1.07	1.11	1.04	1.09	1.00	1.03	0.961			
Neutral (11)	7.00	3.19	5.46	3.65	3.80	3.37	<0.001	<0.001	0.028	-
Surprise (9)	6.93	1.89	5.44	2.31	4.50	2.50	<0.001	<0.001	-	0.001
Sadness (8)	4.10	2.03	3.54	1.70	3.69	2.29	0.192			
Errors of omission	0.12	0.63	0.72	3.72	1.26	9.03	0.010	0.018	-	0.038
Errors of commission	18.20	9.22	22.93	8.33	25.93	12.53	<0.001	<0.001	-	0.001
Mean Reaction Time (sec)	4.42	1.65	6.46	3.05	7.72	4.21	<0.001	<0.001	-	<0.001
Precision of processing	23.31	25.63	3.21	25.72	-10.15	32.77	<0.001	<0.001	0.048	0.001
Processing Speed	3.00	5.27	-1.99	7.99	-7.97	16.62	<0.001	<0.001	0.014	<0.001

Notes: AD = Alzheimer´s disease; aMCI = Amnestic Mild Cognitive Impairment; CAMCOG-R = Revised Cambridge Examination of the Elderly for Mental Disorders. Cognitive part; CDR = Clinical Dementia Rating; GDS = Global Deterioration Scale; GDS-D = Geriatric Depression Scale; H = Healthy older adults; M = Mean; MMSE = Minimental State Examination; n = number of participants; S.D. = Standard Deviation.

Table 2. Pearson´s correlations between Affect-GRADIOR. MMSE. CAMCOG-R and GDS-D scales

	MMSE (n = 212)		CAMCOG-R (n = 70)		GDS-D (n = 93)	
	Score (p)	Processing Speed (p)	Score (p)	Processing Speed (p)	Score (p)	Processing Speed (p)
Happiness	0.228 (0.001)*	0.510 (< 0.001)*	0.017 (0.896)	0.331 (0.009)*	-0.061 (0.501)	-0.031 (0.733)
Disgust	0.192 (0.005)*	0.319 (< 0.001)*	0.217 (0.092)	0.224 (0.083)	-0.148 (0.101)	0.027 (0.761)
Anger	0.167 (0.015)*	0.207 (0.002)*	0.097 (0.456)	0.247 (0.055)	-0.076 (0.400)	-0.159 (0.077)
Fear	0.011 (0.874)	0.178 (0.009)*	0.031 (0.813)	0.200 (0.122)	-0.021 (0.820)	0.032 (0.727)
Neutral	0.208 (0.002)*	0.352 (< 0.001)*	0.514 (<0.001)*	0.465 (< 0.001)*	-0.134 (0.136)	-0.065 (0.471)
Surprise	0.392 (< 0.001)*	0.484 (< 0.001)*	0.227 (0.078)	0.302 (0.018)*	0.127 (0.159	0.015 (0.872)
Sadness	0.092 (0.182)	0.312 (< 0.001)*	0.113 (0.385)	0.131 (0.314)	0.078 (0.385)	0.113 (0.211)
Complete Test (53)	0.487 (< 0.001)*	0.447 (< 0.001)*	0.432 (< 0.001)*	0.251 (0.051)	-0.091 (0.314)	0.013 (0.881)

Notes: * = p < 0.05; CAMCOG-R = Revised Cambridge Examination of the Elderly for Mental Disorders (Cognitive part); CDR = Clinical Dementia Rating; GDS-D = Geriatric Depression Scale; MMSE = Minimental State Examination; n = number of participants assessed; p = p-value for a two-tailed test with a 95% confidence interval.

Table 3. Quantitative Usability data

	H (n = 11)		aMCI (n = 8)		AD (n = 8)			
	M	***SD***	**M**	***SD***	**M**	***SD***	***Chi***	***p***
IT_6	2.09	0.83	2.13	0.84	1.50	0.76	3.062	0.216
IT_7	1.09	0.30	1.38	0.52	1.25	0.46	2.131	0.344
IT_8	1.64	0.81	1.00	0.00	1.25	0.46	5.038	0.081
IT_9	1.09	0.30	1.50	0.54	2.38	1.19	9.596	0.008*
IT_10	1.00	0.00	1.25	0.46	2.38	1.19	12.834	0.002*

Notes: * = post-hoc analysis showed statistically significant differences between AD-aMCI and AD-HC; AD = Alzheimer´s disease; aMCI = Mild Cognitive Impairment; H = Healthy older adults; IT = item of the usability questionnaire (supplemental online material 2); M = Mean; n = number of participants; *SD* = Standard Deviation.

Table 4. Tetrachoric cumulative variance accounted for

	Sums of squared loadings	% of variance	Cumulative %
Neutral	7.934	14.970	14.970
Happiness	6.120	11.548	26.517
Sadness	3.854	7.271	33.788
Disgust	3.902	7.363	41.151
Fear	2.718	5.128	46.279
Surprise	4.546	8.577	54.856
Anger	2.807	5.296	60.152

Table 5. Sensitivities and specificities at optimal cut-off scores for Affect-Gradior main scores.

Group	*Score*	*AUC*	*Cut-off score*	*Sensitivity*	*Specificity*	*AUC 95% CI*
a. HC vs. AD	Correct answers	0.791	28.50	0.761	0.750	0.719-0.863
	Processing speed	0.829	0.85	0.775	0.786	0.765-0.894
b. HC vs. aMCI	Correct answers	0.717	30.50	0.676	0.667	0.629-0.804
	Processing speed	0.748	2.69	0.563	0.877	0.664-0.832
c. aMCI vs. AD	Correct answers	0.642	26.50	0.667	0.643	0.549-0.735
	Processing speed	0.670	-0.17	0.667	0.643	0.581-0.760

Notes: AD = Alzheimer's disease; aMCI = amnestic mild cognitive impairment; AUC = area under the curve; CI = confidence interval; HC = healthy controls.

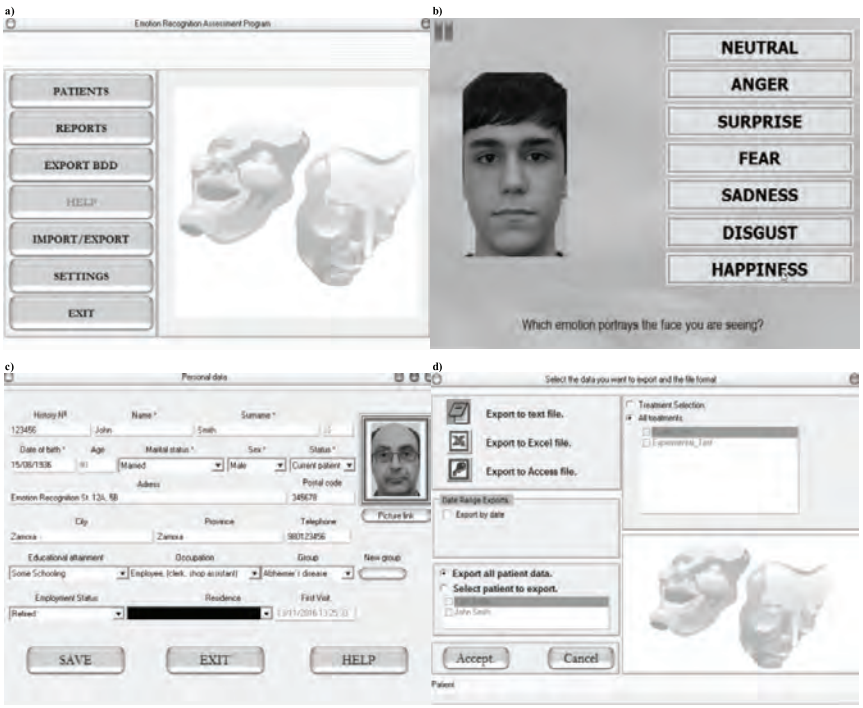

Figure 1. Affect-Gradior Emotion Recognition Test: a) clinician interface; b) patients´ interface; c) clinical records; d) data export.

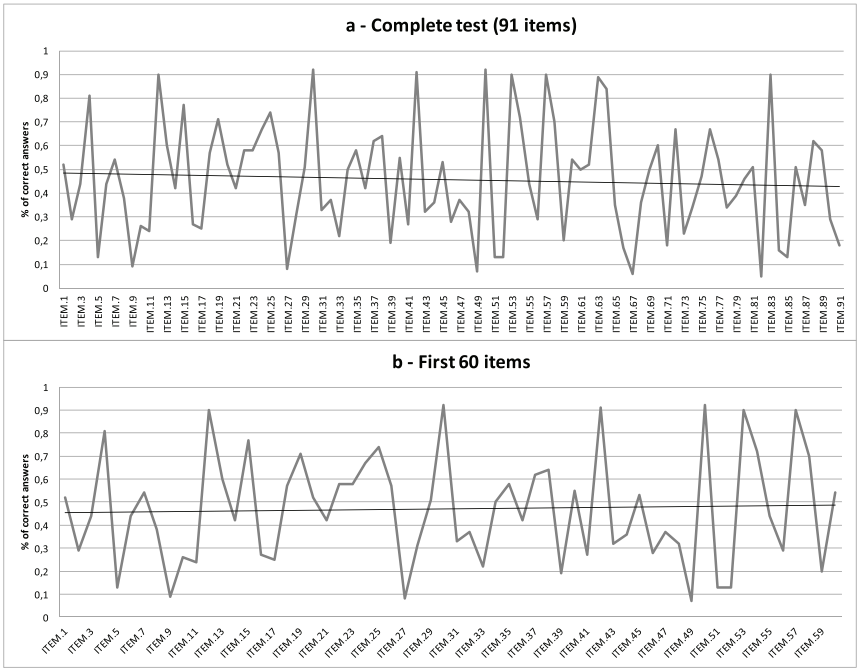

Figure 2. Trend line of the percentage of correct answers along the Affect-Gradior test.

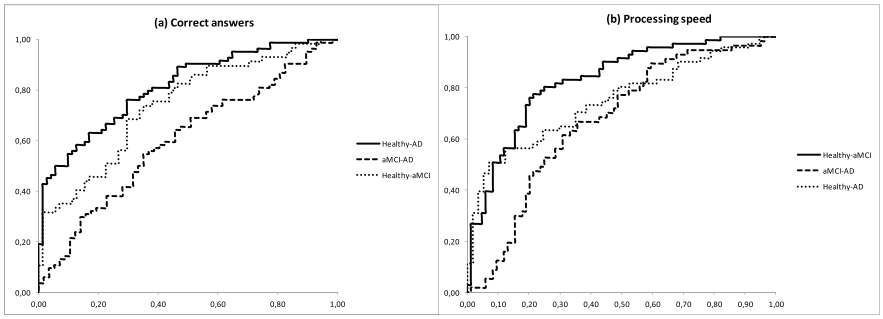

Figure 3. Receiver-operating characteristic (ROC) curve for the comparison of sensitivity to Alzheimer´s Disease (AD), amnestic mild cognitive impairment (aMCI) and healthy controls (HC) in total correct answers (a) and processing speed (b).

Supplemental online material 1

1A - STARD checklist for reporting of studies of diagnostic accuracy

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	5-6
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	6-7
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	6
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	7
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	7
Test methods	7	The reference standard and its rationale.	8-10
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	7-10
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	8-10
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	7
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	7
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	10-11
	13	Methods for calculating test reproducibility, if done.	11-12
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	7
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	Table 1
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is **strongly** recommended).	6
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	N/A
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	Table 1
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	Tables 1-7
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	Figure 3 Table 7 Page 16
	22	How indeterminate results, missing data and outliers of the index tests were handled.	N/A
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	Figure 3 Table 7 Page 16
	24	Estimates of test reproducibility, if done.	13-14
DISCUSSION	25	Discuss the clinical applicability of the study findings.	16-22

1B - STROBE Statement. Checklist of items that should be included in reports of observational studies

YOU MUST CHECK EACH ITEM AS APPROPRIATE FOR YOUR PAPER AND SPECIFY PAGE(S) OF MANUSCRIPT WHERE INFORMATION CAN BE FOUND.

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract ☒
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found ☒ page(s) Page 1
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported ☒ page(s) Pages 2-5
Objectives	3	State specific objectives, including any prespecified hypotheses ☒ page(s) Page 6
Methods		
Study design	4	Present key elements of study design early in the paper ☒ page(s) Methods Page 6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection ☒ page(s) Methods-Procedures page 7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up ☐ page(s) Click here to enter text. *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls ☐ page(s) Click here to enter text. *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants ☒ pages Methods-Participants page 6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed ☐ page(s) Click here to enter text. *Case-control study*—For matched studies, give matching criteria and the number of controls per case ☐ page(s) Click here to enter text.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable ☒ page(s) Diagnostic criteria-Page 6 / outcomes-pages 8-10
Data sources/	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods

measurement		if there is more than one group ☒ page(s) outcome measures Pages 8-10
Bias	9	Describe any efforts to address potential sources of bias ☒ page(s) Multiple regression analysis Page 11
Study size	10	Explain how the study size was arrived at ☒ page(s) Page 7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why ☒ page(s) Results section, pages 11-16
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding ☒ page(s) 10 and 12
		(*b*) Describe any methods used to examine subgroups and interactions ☒ page(s) 10
		(*c*) Explain how missing data were addressed ☐ page(s) N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed **[N/A]** *Case-control study*—If applicable, explain how matching of cases and controls was addressed ☐ page(s) N/A *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy ☐ page(s) N/A
		(*e*) Describe any sensitivity analyses ☒ page(s) ROC and AUC pages 15-16

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed ☐ page(s) N/A
		(b) Give reasons for non-participation at each stage ☒ page(s) N/A
		(c) Consider use of a flow diagram ☒ page(s) N/A
Descriptive data	14*	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders ☒ page(s) Table 1 – page 30
		(b) Indicate number of participants with missing data for each variable of interest ☒ page(s) Table 1 – page 30
		(c) *Cohort study*—Summarise follow-up time (e.g., average and total amount) ☐ page(s) N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time ☐ page(s) N/A

		Case-control study—Report numbers in each exposure category, or summary measures of exposure ☒ page(s) Table 1 – Page 30 / Table 2 – Page 31
		Cross-sectional study—Report numbers of outcome events or summary measures ☒ page(s) N/A
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included ☒ page(s) Table 1 – Page 30
		(*b*) Report category boundaries when continuous variables were categorized ☒ page(s) Table 7 – Page 34
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period ☐ page(s) N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses ☒ page(s) Tables 3-6, pages 33-34
Discussion		
Key results	18	Summarise key results with reference to study objectives ☒ page(s) 16
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias ☒ page(s) 21-22
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence ☒ page(s) 17-22
Generalizability	21	Discuss the generalizability (external validity) of the study results ☒ page(s) 21
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based ☒ page(s) 22

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Supplemental online material 2

Usability questionnaire[1]

	Overall impression
1	How did you find the test?
2	How did you find the test extension?
3	Did you find the instructions clear enough?
4	Were you able to identify the emotions portrayed by the actors?
5	How did you find working with a touchscreen?

	Specific questions for the participant
6	How difficult did you find completing the test?
	(1) very easy (2) easy (3) neither easy nor difficult (4) difficult (5) very difficult
7	How difficult did you find understanding the test?
	(1) very easy (2) easy (3) neither easy nor difficult (4) difficult (5) very difficult
8	Did you have any previous experience with computers?
	(1) no experience (2) some experience (3) wide experience

	Specific questions for the researcher
9	Which level of difficulty had the participant to complete the test?
	(1) very easy (2) easy (3) neither easy nor difficult (4) difficult (5) very difficult
10	Which level of difficulty had the participant to understand the test?
	(1) very easy (2) easy (3) neither easy nor difficult (4) difficult (5) very difficult
	Notes and observations during the assessment

[1] Adapted to this study from: Solís Rodríguez, A. (2014). *Estudio preliminar del cogval-senior, una nueva prueba informatizada para la detección de la demencia Alzheimer en personas mayores.* Doctoral Thesis. University of Salamanca: Salamanca.

Supplemental online material 3

Table S1. Distribution of the discriminatory power of the items

IT_01	IT_02	IT_03	IT_04	IT_05	IT_06	IT_07
IT_08	IT_09	IT_10	IT_11	IT_12	IT_13	IT_14
IT_15	IT_16	IT_17	IT_18	IT_19	IT_20	IT_21
IT_22	IT_23	IT_24	IT_25	IT_26	IT_27	IT_28
IT_29	IT_30	IT_31	IT_32	IT_33	IT_34	IT_35
IT_36	IT_37	IT_38	IT_39	IT_40	IT_41	IT_42
IT_43	IT_44	IT_45	IT_46	IT_47	IT_48	IT_49
IT_50	IT_51	IT_52	IT_53	IT_54	IT_55	IT_56
IT_57	IT_58	IT_59	IT_60	IT_61	IT_62	IT_63
IT_64	IT_65	IT_66	IT_67	IT_68	IT_69	IT_70
IT_71	IT_72	IT_73	IT_74	IT_75	IT_76	IT_77
IT_78	IT_79	IT_80	IT_81	IT_82	IT_83	IT_84
IT_85	IT_86	IT_87	IT_88	IT_89	IT_90	IT_91

Notes: grey items were eliminated due to their low discriminatory power (< 0.03); IT = item.

Supplementary Online Material 4

Table S2. Rotated 7 component solution for Affect-GRADIOR test[a]

	Component								Component						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
NEUTRAL_13_ITEM.86.10	.756							DISGUST_05_ITEM.36.9				.680			
NEUTRAL_04_ITEM.23.4	.708							DISGUST_07_ITEM.56.3				.567			
NEUTRAL_05_ITEM.26.5	.691							DISGUST_11_ITEM.70.5				.539			
NEUTRAL_10_ITEM.61.11	.677							DISGUST_08_ITEM.60.12				.525			
NEUTRAL_08_ITEM.45.1	.672							DISGUST_10_ITEM.66.11				.502			
NEUTRAL_06_ITEM.29.7	.647							DISGUST_09_ITEM.65.7				.494			
NEUTRAL_11_ITEM.68.12	.630							DISGUST_04_ITEM.20.8				.472			
NEUTRAL_09_ITEM.55.2	.606							SADNESS_04_ITEM.21.7					.562		
NEUTRAL_03_ITEM.22.9	.599							SADNESS_02_ITEM.14.6					.530		
NEUTRAL_07_ITEM.31.3	.565							SADNESS_10_ITEM.69.2			.300		.480		
NEUTRAL_12_ITEM.79.13	.528							SADNESS_12_ITEM.80.10					.475		
HAPPINESS_04_ITEM.30.8		.723						SADNESS_08_ITEM.44.13					.459		
HAPPINESS_07_ITEM.53.5		.713						SADNESS_01_ITEM.6.9					.449		
HAPPINESS_05_ITEM.42.3		.693						SADNESS_07_ITEM.40.8					.437		
HAPPINESS_06_ITEM.50.11		.687						SADNESS_06_ITEM.38.3					.418		
HAPPINESS_08_ITEM.57.6		.653						ANGER_12_ITEM.75.7						.685	
HAPPINESS_13_ITEM.83.1		.597						ANGER_05_ITEM.34.1						.587	
HAPPINESS_09_ITEM.63.13		.545						ANGER_07_ITEM.41.6						.502	
HAPPINESS_10_ITEM.64.12		.417	.340					ANGER_02_ITEM.8.10						.431	
SURPRISE_10_ITEM.62.2			.636					ANGER_08_ITEM.43.4						.380	
SURPRISE_08_ITEM.54.8			.607					FEAR_07_ITEM.52.6							.580
SURPRISE_11_ITEM.76.9			.563					FEAR_09_ITEM.67.1							.553
SURPRISE_02_ITEM.19.13			.494					FEAR_04_ITEM.32.2							.508
SURPRISE_03_ITEM.25.3			.481					FEAR_03_ITEM.16.7							.494
SURPRISE_09_ITEM.58.10	.341		.476					FEAR_08_ITEM.59.8							.350
SURPRISE_13_ITEM.89.4			.429				.300								
SURPRISE_12_ITEM.88.7			.414												
SURPRISE_04_ITEM.28.12			.400												

Notes: Principal component analysis with Varimax rotation and Kaiser normalization. Component loadings less than 0.30 were not printed to increase readability. A = The rotation converged in 7 iterations.