Electronic Supplementary Material 1

1. Description of measures used for construct validation (in alphabetical order)

ASEBA Adult Self-Report (ASR; Achenbach & Rescorla, 2003). The ASR is a 123-item self-report questionnaire for adults between 18 and 59 years assessing adaptive functioning (six scales: Friends; Spouse/Partner; Family; Job; Education, Personal Strengths), syndromes (eight scales: Anxious/Depressed; Withdrawn; Somatic Complaints; Thought Problems; Attention Problems; Aggressive Behavior; Rule-breaking Behavior, and Intrusive), DSM-oriented scales (six scales: Depressive Problems; Anxiety Problems; Somatic Problems; Avoidant Personality Problems; Attention Deficit/ Hyperactivity Problems; and Antisocial Personality Problems), and substance use. The majority of the items are presented in a Likert-type format with three response options (not true – somewhat or sometimes true – very true or often true). The ASR is widely used in various domains of psychological research and practice and has good psychometric properties (Achenbach & Rescorla, 2003). We used the official Dutch translation provided by the publisher, available at http://www.aseba.org/ordering/translations.html.

Big Five Inventory (BFI; Plaisant, Courtois, Réveillère, Mendelsohn, & John, 2010). The 44-item BFI is one of the most widely used self-report questionnaires assessing the five personality dimensions Extraversion, Openness, Conscientiousness, Neuroticism, and Agreeableness. Participants state their agreement to various characteristics (e.g., "I am someone who is talkative") on a five-point Likert scale. We used the French version validated by Plaisant et al. (2010).

Emotion Recognition Index (ERI; Scherer & Scherer, 2011). The ERI is a 10-minute test to measure emotion recognition ability (ERA). It consists of a facial and a vocal subtest; each one

measuring accuracy in detecting sad, fearful, angry, happy, and neutral expressions. The facial subtest includes 30 pictures of posed expressions from the Pictures of Facial Affect set (Ekman & Friesen, 1976) that are each presented for 3 seconds. The vocal subtest is composed of 30 brief recordings produced by radio actors (Banse & Scherer, 1996). In these recordings, the verbal content is one of two pseudo-linguistic sentences, i.e., sentences that resemble natural speech but are meaningless. In both subtests, after each portrayal participants are asked to choose (without time limitation) which emotion had been expressed. A total score is calculated from the number of items in which the participant's response matched the target emotion. The Dutch translation was provided by Prof. Johnny Fontaine (second author).

Interpersonal Reactivity Index (IRI; Davis, 1983). The IRI is a widely used 28-item self-report questionnaire and measures cognitive and affective dimensions of empathy. It consists of four subscales: Perspective-taking, Fantasy, Empathic Concern, and Personal Distress. Participants indicate to what extent various statements (e.g., "I often have tender concerned feelings for people less fortunate than me") are true about them on a 5-point scale ranging from "not true about me" to "extremely true about me"). Perspective-taking is theorized to capture a more cognitive facet of empathy, whereas empathic concern and fantasy reflect a more affective facet of empathy. Personal distress captures a maladaptive facet of empathy. We used the French version developed by Guttman and Laporte (2002).

Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000). The JACBART is a test to measure ERA and consists of 56 pictures of emotion expressions produced by Japanese and Caucasian individuals, each representing one of seven emotions (surprise, sadness, anger, happiness, fear, disgust, and contempt). Each picture is displayed for 200 ms between two 500 ms presentations of the same individual with a neutral facial expression. After each of these 56 stimuli, participants choose which of the seven emotions

had been shown in the brief emotion expression. A total ERA score is calculated as the percentage of correct responses. Here, we used a version that was programmed and validated in French by Bänziger et al. (2009) according to the information provided in Matsumoto et al. (2000) with the original stimuli, as the JACBART is no longer available from the test authors.

Multimodal Emotion Recognition Test (MERT; Bänziger, Grandjean, & Scherer, 2009). The MERT is a test to measure ERA and consists of 30 video clips of actors (three for each of 10 emotions: Irritation, anger, anxiety, fear, happiness, elated joy, disgust, contempt, sadness, despair) that are presented in four modalities (still picture, video only, audio only, audio with video), yielding a total of 120 items. These recordings were taken from a corpus developed by Banse and Scherer (1996) and a standard pseudo-linguistic sentence is used as verbal content. After each of the 120 stimuli, participants are asked to choose which of the 10 emotions had been expressed by the actor. We calculated a total ERA score as the percentage of correctly recognized items. We used the original French version developed by Bänziger et al. (2009), and the Dutch translation was provided by Prof. Johnny Fontaine (Ghent University).

Mini Profile of Nonverbal Sensitivity (MiniPONS; Bänziger, Scherer, Hall, & Rosenthal, 2011). The MiniPONS is a test measuring interpersonal sensitivity and consists of 64 portrayals in which a young woman displays different affective states in one of six modalities (face, two voice versions, body, and combinations of the voice versions with face). After each portrayal, participants chose which one out of two affective states (e.g., asking forgiveness vs. threatening someone) had been expressed by the actor. Interpersonal sensitivity is calculated as the total percentage of correct choices. We used the original French version validated by Bänziger et al. (2011) and the Dutch translation was provided by Prof. Johnny Fontaine (Ghent University).

Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso, & Sitarenios, 2003). The MSCEIT is a performance-based test of emotional intelligence. It

consists of 141 items that measure four branches of EI (Perceiving Emotions, Facilitating Thought, Understanding Emotions, Managing Emotions) using eight different tasks such as identifying emotions from abstract pictures, evaluating how moods impact thinking, and indicating the effectiveness of different solutions to problems. The MSCEIT yields scores for each branch and an overall EI score. We used the official French online version available from the test publisher, Multi-Health Systems (participants select French as their language after logging in to the assessment tool).

NV5R (Thiébaut & Bidan-Fortier, 2003). This battery of tests measures cognitive aptitudes in young adults and has been normed and validated in French. We administered three subtests of the NV5R, namely general reasoning (inductive and deductive reasoning with numeric, spatial, and lexical content), attention (ability to maintain mental effort and perception speed), and vocabulary (knowledge of French, identifying synonyms and antonyms). As specified by the test authors, these subtests were administered by paper and pencil with a time limit of 20, 3, and 4 minutes, respectively. For each subtest, we used a total score calculated as the percentage of correct responses.

Rosenberg Self-Esteem Scale (RSES; Rosenberg; 1965). The RSES is a widely used questionnaire that assesses global self-esteem by measuring both positive and negative feelings about the self. It contains ten items that respondents answer on a four-point Likert scale from "strongly agree" to "strongly disagree". A mean average score is calculated to reflect global self-esteem. Here, we used the Dutch version validated by Franck, De Raedt, Barbez, and Rosseel (2008).

State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The STAI differentiates between the temporary condition of state anxiety and the longstanding quality of trait anxiety. Here, we only administered the trait part of the

questionnaire. In 20 items, participants indicate on a four-point Likert scale ranging from "almost never" to "almost always" how often they experience various feelings about themselves. Trait anxiety is calculated as the average score of all items. Here, we used the French version developed and validated by Gauthier and Bouchard (1993).

State-Trait Anger Expression Inventory (STAXI; Spielberger, 1988). The STAXI is a 44-item self-report questionnaire consisting of three parts which measure state anger, trait anger, and anger expressivity and control. Here, we administered the trait anger and the expressivity/ control scales which consist of 34 items in total. On these scales, participants indicate on a four-point Likert scale ranging from "almost never" to "almost always" how often they show certain behaviors. Here, we used the French translation validated by Borteyrou, Bruchon-Schweitzer, & Spielberger (2008).

Situational Test of Emotional Understanding (STEU; MacCann & Roberts, 2008). The STEU is a performance-based test measuring knowledge of emotion antecendents and features. It consists of short written vignettes for 14 different emotions, for each of which participants are asked to choose the emotions they represent (or vice versa) out of five response options. Each situation is presented in a context-reduced, a personal-life, and a work-life context version, yielding a total of 42 items. The correct answers are based on theoretical grounds and responses are scored as correct or incorrect accordingly. In Studies 4 and 5 (FR2 and FR3), we used the 25-item short version of the STEU (MacCann, C., & Roberts, R. D., 2012: The Brief assessment of emotional intelligence: Short forms of the situational test of emotional understanding (STEU) and situational test of emotion management (STEM). Technical Report, Educational Testing Service). This STEU version was translated by collaborators from the Swiss Center for Affective Sciences (University of Geneva, Switzerland) into French using the translation/ back-translation

procedure. In Study 2 (DU2) we used the full 42-item version which was translated and validated by Libbrecht and Lievens (2012).

Situational Test of Emotion Management (STEM; MacCann & Roberts, 2008). The STEM is a performance-based test measuring the ability to regulate emotions in other people. Participants are presented with 44 short vignettes and are asked to choose from four responses the most effective course of action. The STEM is scored according to expert ratings, where each response is scored with a weight ranging from zero to one which has been emiprically derived from an expert sample. Here, we used the 30-item version of the STEM which was translated and validated by Libbrecht and Lievens (2012).

Toronto Alexithymia Scale (TAS; Bagby, Parker, & Taylor, 1994). The TAS is a 20-item self-report questionnaire that measures three facets of alexithymia, namely difficulty in identifying one's feelings, difficulty in describing one's feelings, and externally-oriented thinking. Items are completed using a 5-point Likert scale ranging from "strongly disagree" to "strongly agree". Here, we used the Dutch version validated by De Gucht, Fontaine, and Fischler (2004).

Trait Emotional Intelligence Questionnaire (TEIQue, Petrides, 2009). The TEIQue is a self-report questionnaire measuring four broad factors of trait emotional intelligence (well-being, self-control, emotionality, and sociability) and 15 more specific facets. Participants are asked to state their agreement with various statements on a 7-point Likert scale ranging from "disagree completely" to "agree completely". Here, we used the TEIQue short form that consists of 30 items. The French version was validated by Mikolajczak, Luminet, Leroy, and Roy (2007). The Dutch version was validated by Mavrovelli, Petrides, Rieffe, and Bakker (2007).

2. Analysis of measurement properties of the French and Dutch GERT using Item Response Theory

Method

In order to examine the difficulty and measurement precision of the Dutch and French versions of the GERT, we analyzed participants' scored responses on the 83 items (1= correct, 0= incorrect) using the Rasch model (Rasch, 1993). The Rasch model is a simple model in Item Response Theory (IRT) that can be used for developing and evaluating scales with binary items. It offers certain advantages over tests developed with Classical Test Theory. In particular, it provides the statistical basis for concluding that it is the same latent trait or ability (here, ERA) that is involved in solving each item regardless of its difficulty or emotion category. As such, in a Rasch-homogeneous test, the mean or sum score of a participant's correct responses is a sufficient estimate of a person's ability.

The basic idea of the Rasch model is that the probability of correctly responding to an item depends on a person's ability (here: ERA) and the difficulty of an item, which are both located on the same latent scale θ. For any item, a person with a higher ability is assumed to have a higher probability of solving it than a person with a lower ability. In addition, items with a lower difficulty have a higher probability of being solved than items with a higher difficulty. These theoretically postulated relationships can be displayed for each item with an Item Characteristic Curve (ICC). The fit of the Rasch model can be assessed by inspecting how much the observed ICCs differs from the ICCs that are theoretically expected from the Rasch model. We assessed model fit using the weighted-fit or "Infit" and unweighted-fit or "Outfit" index for each item provided by the eRm package in R (Mair & Hatzinger, 2007). Values between .80 and 1.20 are usually considered an "indication of useful fit" (Wright, Linacre, Gustafson, & Martin-Lof, 1994), with 1.00 representing perfect fit. After establishing model fit, we evaluated test and item difficulties as well as measurement precision by inspecting the so-called *Test Information Curve* (TIC) which shows the range of the latent dimension θ in which the test discriminates best

among individuals, the *standard error of measurement* (SDEM) in the range of θ where the ability scores of most participants were located. Furthermore, we calculated the total *test score reliability* from the IRT parameters following the method proposed by Dimitrov (2003) for binary data.

Results

Descriptive statistics for the full GERT and the emotion subscales are shown in Table 1. Results showed that Infits were in the range of useful fit for all items in both the Dutch (.89 to 1.10) and the French (.80 to 1.13) GERT versions. Outfits ranged from .84 to 1.13 in the Dutch data and from .67 to 1.34 in the French data, with three out of the 83 French GERT items (irr44, rel66, and sur74) not meeting the criterion of useful fit. Overall, these results suggest a good fit of the Rasch model to the data for both languages, replicating the results for the German GERT.

We then evaluated the item difficulty distribution on the θ dimension in comparison to participants' ability parameters, the Test Information Curves, and the Standard errors of measurement (Figures 1 and 2). Figure 1 shows that for the Dutch GERT, the distribution of the item difficulties matches the distribution of the person parameters quite well (mean ability estimate =.24, *SD*=.61). Furthermore, the Test Information Curve and the Standard Errors of Measurement indicate that the GERT provides most measurement information and the highest precision in the θ range on which most Dutch participants scored, showing that the GERT has an appropriate difficulty level for this sample. Test score reliability in the Dutch sample was .93. Figure 2 shows that the French sample on average achieved higher scores than the Dutch sample (mean ability estimate= .87, SD= .58) and that a large part of the GERT items were somewhat too easy for the French participants, similarly to the results of the German GERT (Schlegel et al., 2014). However, for 95% of the person ability parameters of the French sample, the Standard Error of Measurement was still acceptable, ranging from .24 to .32. The 95% confidence interval

for an ability estimate of -0.31 (the 2.5th percentile of the population) [-.78, .16] is thus by only .32 (about ½ standard deviation) larger than the confidence interval for an ability estimate of 2.0 (the 97.5th percentile) [1.37, 2.63]. Thus, measurement precision was still comparatively good for the large majority of the French sample. Test score reliability (Dimitrov, 2003) in the French sample was the same as in the Dutch sample (.93). Overall, these results support the good psychometric quality and high measurement precision of the GERT in French and Dutch.

References

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*(5), 691–704.

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A Short Multichannel Version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior*, *35*, 189–204.

Borteyrou, X., Bruchon-Schweitzer, M., & Spielberger, C. D. (2008). Une adaptation française du STAXI-2, inventaire de colère-trait et de colère-état de C.D. Spielberger. *L'Encéphale*, *34*(3), 249–255.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126.

De Gucht, V., Fontaine, J., & Fischler, B. (2004). Temporal stability and differential relationships with neuroticism and extraversion of the three subscales of the 20-item Toronto Alexithymia Scale in clinical and nonclinical samples. *Journal of Psychosomatic Research, 57*(1), 25–33.

Dimitrov, D. M. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement*, *27*(6), 440–458.

Franck, E., De Raedt, R., Barbez, C., & Rosseel, Y. (2008). Psychometric properties of the Dutch Rosenberg self-esteem scale. *Psychologica Belgica*, *48*(1), 25-35.

Gauthier, J., & Bouchard, S. (1993). Adaptation canadienne-française de la forme revisée du State-Trait Anxiety Inventory de Spielberger, *Revue Canadienne des Sciences du Comportement, 25*(4), 559-578.

Guttman, H., & Laporte, L. (2002). Alexithymia, empathy, and psychological symptoms in a family context. *Comprehensive Psychiatry, 43(6)*, 448–455.

Hall, J. A. (2001). The PONS Test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity: Theory and Measurement* (pp. 145–162). Lawrence Erlbaum Associates.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Libbrecht, N., & Lievens, F. (2012). Validity evidence for the situational judgment test paradigm in emotional intelligence measurement. *International Journal of Psychology, 47*(6), 438–447.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*(4), 540–551.

Mair, P., & Hatzinger, R. (2007). *Extended Rasch modeling: The eRm package for the application of IRT models in R*. WU Vienna University of Economics and Business.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., … Goh, A. (2000). A New Test to Measure Emotion Recognition Ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, *24*(3), 179–209.

Mavroveli, S, Petrides, K.V., Rieffe, C. & Bakker, F. (2007). Trait emotional intelligence, psychological well-being, and peer-rated social competence in adolescence. *British Journal of Developmental Psychology, 25*, 263-275.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*(1), 97–105.

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric properties of the Trait Emotional Intelligence Questionnaire: Factor structure, reliability, construct, and incremental validity in a French-speaking population. *Journal of Personality Assessment, 88*(3), 338–353.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*(1), 9–35.

Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire. In C. Stough, D. H. Saklofske, and J. D. Parker, *Advances in the Assessment of Emotional Intelligence* (pp. 85-101). New York: Springer.

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., & John, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Annales Médico-Psychologiques, Revue Psychiatrique*, *168*(2), 97–106.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests.* MESA Press.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, *8*(3), 370.

Table 1

Descriptive Statistics of the GERT Subscales in the Dutch and French Samples

	Dutch GERT (*N*=966)		French GERT (*N*=317)	
emotion	*M (SD)*	H_u *(SD)*	*M (SD)*	H_u *(SD)*
amusement	.57 (.27)	.47 (.28)	.80 (.23)	.70 (.26)
anger	.69 (.24)	.53 (.24)	.76 (.20)	.65 (.21)
disgust	.54 (.23)	.43 (.24)	.71 (.20)	.63 (.22)
despair	.47 (.25)	.22 (.18)	.47 (.26)	.27 (.21)
pride	.47 (.24)	.34 (.23)	.58 (.24)	.47 (.25)
anxiety	.58 (.24)	.26 (.17)	.69 (.23)	.39 (.21)
interest	.53 (.24)	.36 (.22)	.60 (.24)	.44 (.23)
irritation	.50 (.25)	.28 (.20)	.68 (.28)	.43 (.24)
joy	.56 (.24)	.33 (.21)	.71 (.23)	.50 (.22)
fear	.47 (.25)	.31 (.23)	.62 (.24)	.49 (.25)
pleasure	.57 (.30)	.39 (.28)	.75 (.23)	.61 (.27)
relief	.83 (.19)	.57 (.23)	.85 (.17)	.71 (.22)
surprise	.42 (.23)	.25 (.18)	.60 (.22)	.36 (.18)
sadness	.50 (.25)	.36 (.24)	.57 (.27)	.42 (.26)
total	.55 (.12)	.36 (.13)	.67 (.11)	.50 (.13)

Note. H_u= unbiased hit rate (Wagner, 1993).

Table 2

Correlation Matrix and Cronbach's Alpha of All Tests and Questionnaires in Study 1 (DU1, N=177)

	α	MiniPONS	GERT	MERT
MiniPONS	.55			
GERT	.84	.19		
Multimodal Emotion Recognition Index (MERT)	.82	.24	.56***	
Emotion Recognition Index (ERI)	.60	.26*	.29**	.51***

Table 3

Correlation Matrix and Cronbach's Alpha of All Tests and Questionnaires in Study 2 (DU2; N=789)

	α	GERT	1	2	3	4	5	6	7	8
(1) Situational Test of Emotional Understanding (STEU)	.56	.46**								
(2) Situational Test of Emotion Management (STEM)	.72	.36**	.38**							
(3) ASR social adjustment	.80	.18**	-.01	.01						
(4) ASR personal strength	.58	.10**	.15**	.18**	.10**					
(5) ASR depressive symptoms	.84	.09*	.03	-.02	-.26**	-.33**				
(6) ASR anxiety symptoms	.77	.07	.06	.03	-.27**	-.22**	.75**			
(7) ASR withdrawal symptoms	.84	.04	.03	-.06	-.42**	-.28**	.61**	.48**		
(8) ASR somatic problems	.78	.07	.02	.01	-.21**	-.19**	.71**	.62**	.41**	
(9) ASR thought problems	.66	.06	.09*	-.08*	-.16**	-.20**	.57**	.52**	.52**	.48**
(10) ASR attention problems	.84	.19**	.08*	-.04	-.13**	-.32**	.70**	.51**	.48**	.48**
(11) ASR aggressive behavior	.81	.11**	0.03	-.10*	-.24**	-.22**	.64**	.57**	.55**	.49**
(12) ASR rule-breaking behavior	.64	.07	-.02	-.15**	-.04	-.13**	.33**	.21**	.32**	.22**
(13) TEIQue wellbeing	.81	.05	.09*	.12**	.28**	.40**	-.59**	-.54**	-.49**	-.40**
(14) TEIQue self-control	.64	.04	.12**	.13**	.08*	.34**	-.50**	-.50**	-.30**	-.39**
(15) TEIQue emotionality	.70	.19**	.19**	.23**	.25**	.30**	-.24**	-.17**	-.41**	-.12**
(16) TEIQue sociability	.64	.11**	.11**	.07	.23**	.32**	-.43**	-.44**	-.36**	-.31**
(17) Rosenberg self-esteem scale	.89	-.04	.03	.11**	.19**	.42**	-.68**	-.56**	-.47**	-.47**
(18) Toronto Alexithymia Scale	.67	-.20**	-.23**	-.23**	-.20**	-.39**	.35**	.27**	.38**	.25**

Table 3, continued

	α	9	10	11	12	13	14	15	16	17
(10) ASR attention problems	.84	.55**								
(11) ASR aggressive behavior	.81	.57**	.60**							
(12) ASR rule-breaking behavior	.64	.44**	.48**	.46**						
(13) TEIQue wellbeing	.81	-.35**	-.41**	-.46**	-.15**					
(14) TEIQue self-control	.64	-.36**	-.46**	-.54**	-.23**	.57**				
(15) TEIQue emotionality	.70	-.22**	-.28**	-.35**	-.20**	.49**	.45**			
(16) TEIQue sociability	.64	-.25**	-.33**	-.27**	-.012	.52**	.49**	.41**		
(17) Rosenberg self-esteem scale	.89	-.43**	-.52**	-.53**	-.24**	.67**	.55**	.34**	.50**	
(18) Toronto Alexithymia Scale	.67	.30**	.33**	.39**	.19**	-.44**	-.48**	-.69**	-.41**	-.46**

Note. ASR= Adult Self Report (ASEBA), TEIQue= Trait Emotional Intelligence Questionnaire

Table 4

Correlation Matrix of All Tests and Questionnaires in Study 3 (FR1, N=131)

	α	GERT	1	2	3	4	5	6	7	8	9	10	11	12	13	14
(1) MERT	.67	.61***														
(2) MiniPONS	.48	.43***	.62***													
(3) JACBART	.86	.49***	.44***	.40***												
(4) Emotion Perception (MSCEIT)	.91	.31***	.22*	.29**	.16											
(5) Emotion Facilitation (MSCEIT)	.71	.40***	.29**	.23*	.27**	.49***										
(6) Emotional Understanding (MSCEIT)	.61	.36***	.42***	.34***	.25**	.29**	.41***									
(7) Emotion Management (MSCEIT)	.70	.41***	.36***	.23**	.24**	.33**	.51***	.44***								
(8) General Reasoning (NV5-R)	.81	.37***	.33***	.28**	.34***	.09	.24**	.45***	.26**							
(9) Attention (NV5-R)	.93	.29**	.30**	.31**	.25**	.12	.10	.09	.04	.25**						
(10) Vocabulary (NV5-R)	.93	.36***	.38***	.40***	.43***	.09	.20**	.30**	.18*	.47***	.56***					
(11) Extraversion	.85	.11	.15	.04	.20*	.19*	.13	.11	.39***	-.04	.09	.10				
(12) Conscientiousness	.80	.00	.28**	.11	-.02	-.12	-.03	.08	.14	.01	.12	.17	.15			
(13) Neuroticism	.80	.30**	.14	.12	.12	.05	-.09	.01	-.17	-.03	.09	.20*	-.10	-.20*		
(14) Openness	.76	.32***	.25**	.13	.20*	.21*	.18*	.22*	.35***	.19*	.04	.16	.31***	.08	.01	
(15) Agreeableness	.67	.11	.19*	.06	.18*	-.02	.15**	.24**	.32***	.12	.12	.09	.15	.42***	-.35***	.14

Table 5

Correlation Matrix of All Tests and Questionnaires in Study 4 (FR2, N=117)

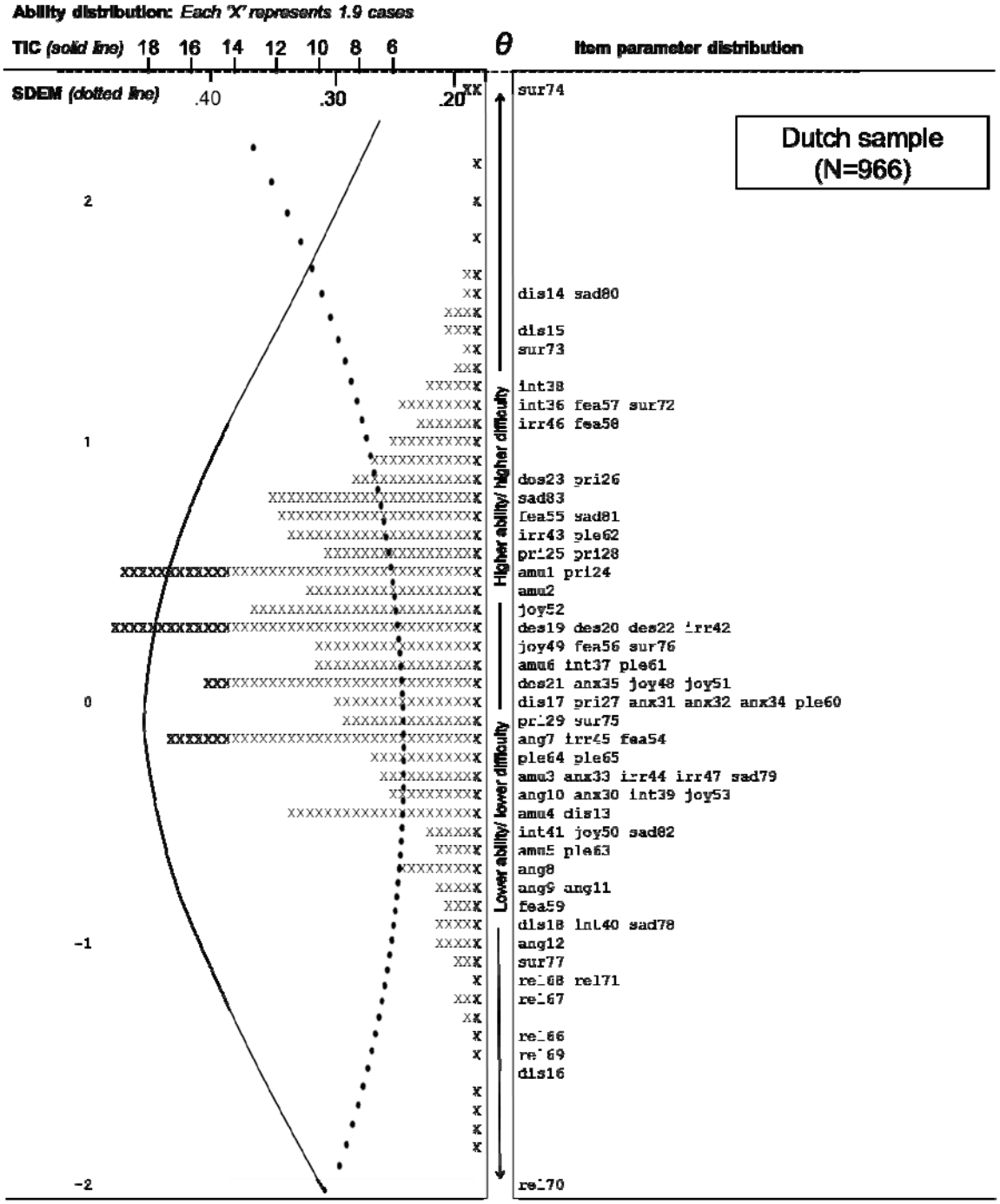
	α	GERT	STEU	wellbeing	self-control	emotionality
Situational Test of Emotional Understanding (STEU)	.78	.50**				
TEIQue wellbeing	.78	-.03	.01			
TEIQue self-control	.57	-.10	.01	.44**		
TEIQue emotionality	.69	.02	-.09	.35**	.09	
TEIQue sociability	.64	.15	.07	.32**	.31**	.14

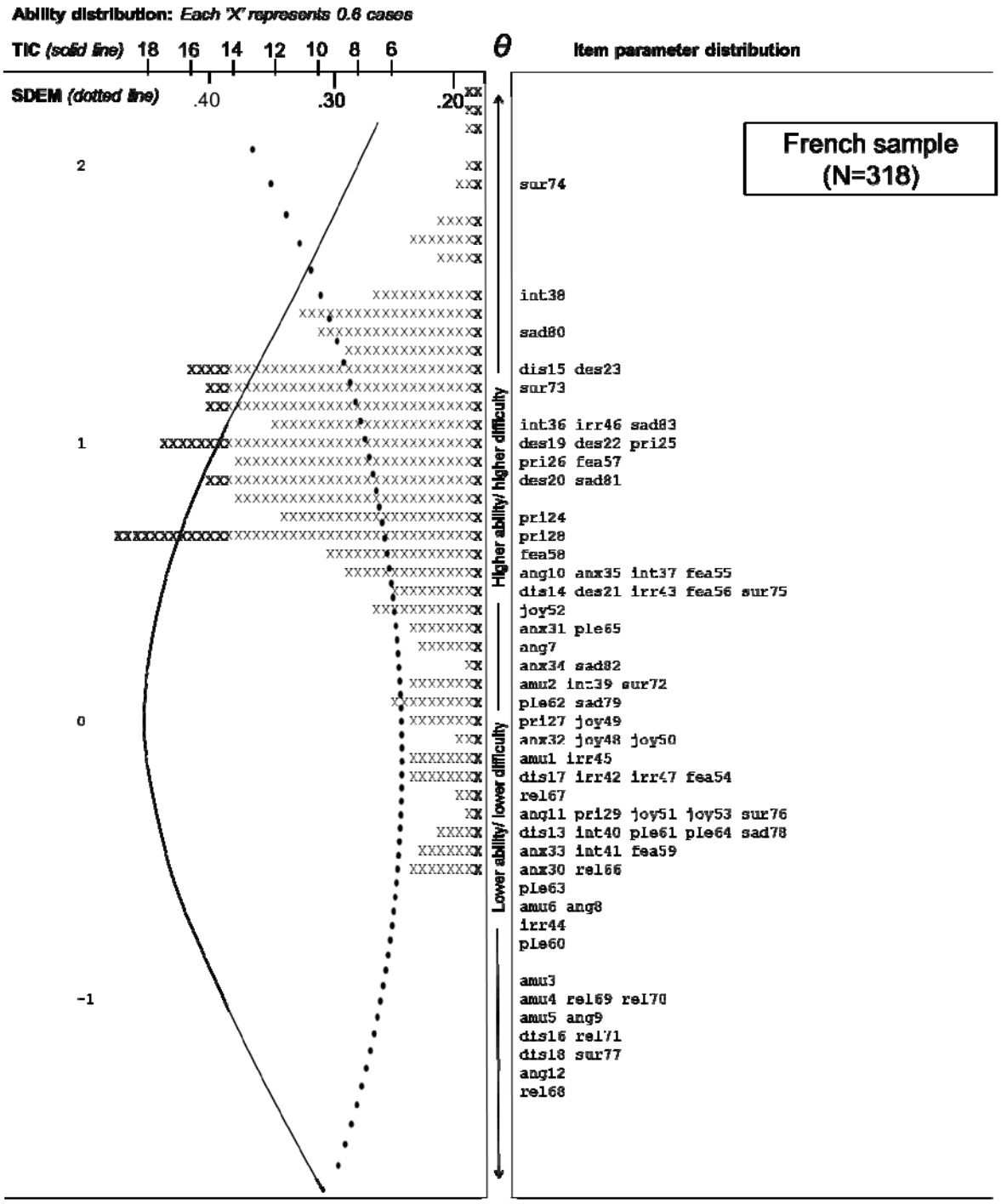
Table 6

Correlation Matrix of All Tests and Questionnaires in Study 5 (FR3, N=70)

	α	GERT	1	2	3	4	5	6	7
(1) Situational Test of Emotional Understanding (STEU)	.67	.59***							
(2) Perspective taking (IRI)	.80	.39**	.23						
(3) Fantasy (IRI)	.80	.45***	.55***	.19					
(4) Empathic Concern (IRI)	.74	.30*	.28*	.62***	.18				
(5) Personal Distress (IRI)	.73	.13	.16	-.05	.28*	.15			
(6) Trait Anxiety (STAI)	.93	-.26*	-.23	-.09	-.12	-.02	.20		
(7) Trait Anger (STAXI)	.81	-.25*	-.22	-.31*	.01	-.09	.43***	.11	
(8) Anger expression and control (STAXI)	.76	-.02	-.19	-.23	.01	-.09	.47***	.32**	.54***

Figures 1 and 2. Wright Maps displaying the distribution of the ability estimates, the Test Information Curve (TIC), and the Standard Error of Measurement (SDEM) on the left side and the item difficulties on the right side of the latent dimension θ for the Dutch and French samples.





Note. The labels on the right side of each figure represent the item numbers (1 to 83) and the respective emotion category. Amu=amusement, ang=anger, dis=disgust, des=despair, pri=pride, anx=anxiety, int=interest, irr=irritation, fea=fear, ple=pleasure, rel=relief, sur=surprise, sad=sadness. The mean of the item difficulties was fixed to zero.