J Nonverbal Behav (2011) 35:189–204
DOI 10.1007/s10919-011-0108-3

ORIGINAL PAPER

Introducing the MiniPONS: A Short Multichannel Version of the Profile of Nonverbal Sensitivity (PONS)

Tanja Bänziger · Klaus R. Scherer · Judith A. Hall · Robert Rosenthal

Published online: 13 April 2011

Abstract Despite extensive research activity on the recognition of emotional expression, there are only few validated tests of individual differences in this competence (generally considered as part of nonverbal sensitivity and emotional intelligence). This paper reports the development of a short, multichannel, version (MiniPONS) of the established Profile of Nonverbal Sensitivity (PONS) test. The full test has been extensively validated in many different cultures, showing substantial correlations with a large range of outcome variables. The short multichannel version (64 items) described here correlates very highly with the full version and shows reasonable construct validity through significant correlations with other tests of emotion recognition ability. Based on these results, the role of nonverbal sensitivity as part of a latent trait of emotional competence is discussed and the MiniPONS is suggested as a convenient method to perform a rapid screening of this central socio-emotional competence.

Keywords Nonverbal sensitivity · Emotional competence · Emotional intelligence · Assessment

T. Bänziger · K. R. Scherer (✉)
Swiss Centre for Affective Sciences, University of Geneva, rue des Battoirs 7, 1205 Geneva, Switzerland
e-mail: Klaus.Scherer@unige.ch

J. A. Hall
Northeastern University, Boston, MA, USA

R. Rosenthal
University of California, Riverside, CA, USA

Present Address:
T. Bänziger
Department of Psychology, Uppsala University, Box 1225, 751 42 Uppsala, Sweden

Introducing the MiniPONS

A Short Multichannel Version of the Profile of Nonverbal Sensitivity (PONS)

This article describes the development and validation of a short version of the established Profile of Nonverbal Sensitivity (PONS), developed by Rosenthal and collaborators (Rosenthal et al. 1979). The PONS was designed to measure individual differences in the ability to recognize emotions, interpersonal attitudes, and communicative intentions in different nonverbal channels or modalities.

The PONS test consists of 220 excerpts of 2-s duration extracted from longer portrayals by a young woman of 20 different interpersonal situations varying widely with respect to their emotional quality. The 20 situations were designed to represent a 2×2 design combining valence and dominance, with 5 items in each of the quadrants formed by the crossing of these two factors: positive-submissive (e.g., expressing gratitude), positive-dominant (e.g., expressing motherly love), negative-submissive (e.g., asking forgiveness), and negative dominant (e.g., threatening someone). The audiovisual recordings of these scenes were edited in such a way as to produce stimuli for 11 channels: full body (head to knees of the standing actor, no sound), face only (filling the screen, no sound), body only (neck to knees, no sound), and the two audio channels of low-pass filtered voice and random-spliced voice (for the details of these procedures masking verbal content and retaining intonation and fluency or voice quality respectively, see Juslin and Scherer 2005; Scherer et al. 1985), as well as all possible combinations of these visual and audio cues. Test-takers respond to the test items using a multiple-choice answer sheet on which, for each item, the choice is between the correct answer and an alternative, randomly selected from among the remaining 19 situations.

Given its length (220 items), the full PONS shows high internal consistency (KR-20 of .86; see Rosenthal et al. 1979) despite low item intercorrelations. In addition, the PONS has shown remarkable predictive validity. In the original PONS monograph Rosenthal et al. reported quite a number of significant results; to name a few, high scorers had healthier, well-adjusted personalities, were rated as more interpersonally sensitive by peers and supervisors, were more democratic as teachers, and were rated as better in their job performance as clinicians and teachers. The PONS was tested in many different cultures and accuracy rates were similar to the U.S. rates depending on the degree of similarity in terms of modernization. Performance generally improved from middle childhood to college, psychiatrically treated samples showed inferior performance to non-psychiatric samples, and married women with toddler-age children scored higher than matched women without children (having to read the nonverbal cues of a preverbal child was predicted to increase the parent's skill; for other evidence supporting an impact of formative influences on nonverbal sensitivity, see Hall et al. 2009). Consistently, over many studies, the PONS revealed gender differences with women scoring on average two percentage points higher (a point-biserial correlation of .20 or Cohen's d of about .40), an effect size identical to that shown in the literature on judging nonverbal affective cues excluding the PONS (Hall 1978, 1984).

Among findings published subsequent to those described in the PONS's initial monograph (Rosenthal et al. 1979) are the following. Among physicians, those scoring higher on the PONS had more satisfied patients (DiMatteo et al. 1979) and were more vigilant for cues of anxiety and depression in their patients (Robbins et al. 1994). Among college samples, high scorers learned more in an interpersonal learning situation (Bernieri 1991), possessed more accurate knowledge of differences in male vs. female behaviors (Hall and

Carter 1999), were seen by friends as warm and likely to be turned to for advice and reassurance (Funder and Harris 1986), had lower depressive tendencies (Ambady and Gray 2002), possessed higher levels of factual knowledge about the uses and correlates of nonverbal cues (Rosip and Hall 2004), were rated by students and observers as more effective music teachers (Kurkul 2007), and were more accurate in judging extraversion and positive affect of interaction partners (Ambady et al. 1995). In addition, adults who scored higher on the PONS in a longitudinal study had easier temperaments as children and experienced more parental harmony (Hodgins and Koestner 1993). For readers interested in deficits of low scorers, many of the statements above can simply be reversed, as the findings are generally based on correlations. For example: "high scorers … had lower depressive tendencies" can be taken to mean that low scorers tend to have higher depressive tendencies.

While the PONS has been criticized for using a single encoder, it is still the only test that has systematically manipulated different nonverbal channels or modalities for comparable stimuli, demonstrating interesting individual differences in the ability to use different nonverbal cues for the inference of emotion, attitudes, and interpersonal stances (Rosenthal et al. 1979). In comparison, most of the relatively few tests of emotion recognition focus on one single modality (but see the Multimodal Emotion Recognition Test—MERT—Bänziger et al. 2009). One of the drawbacks of the test is its length, requiring about an hour of highly concentrated and somewhat repetitive work. Some short forms, described in Rosenthal et al., have been developed by selecting items from the full version. The two most frequently used are the 40-item Audio PONS test, containing the 20 content-filtered and 20 random-spliced items from the full PONS, and the 40-item Video PONS test, containing the 20 face-only and 20 body-only items from the full PONS. As one might expect on the basis of the low inter-item correlations (see above), the internal consistency of these short forms is generally very low (Hall 2001). However, in several studies, sizeable correlations with background or outcome variables have been found. Hall (2001) has suggested that this might be due to the fact that the standard psychometric model, assuming random error to cause low internal consistency, may not be applicable to nonverbal sensitivity tests. Rather, it may be that such tests actually gain validity by including items that represent a number of different, albeit related skills (i.e., differential recognition ability for different emotions or different channels) rather than assuming essentially a replication of item content as in classic psychometric theory (see also Bollen and Lennox 1991).

If this is indeed the case, one would assume that it would be preferable to develop a short *multichannel* form (MiniPONS) of the PONS that contains most of the channels and channel combinations of the original PONS, sampling a large number of potentially different channel-specific decoding skills (rather than using audio or video only items as in the existing short forms). This was the central aim of the current work: Attempting to reduce the length of the test to a duration of less than 15 min but maintaining a systematic sampling of major channels and channel combinations. As a criterion for item selection, we used a fixed number of items for a representative number of design conditions (channels by quadrants), selecting items with average levels of difficulty.

An additional aim was to produce a computerized version of the short test using digitized video and audio presentations as well as programmed response handling and data storage, thus allowing research via web administration. Below we describe two studies: Study 1 which served to develop the short version and examine its relationship to the full PONS as well as to evaluate construct validity by correlating the test scores with those of a number of established test in the area, and Study 2 which served to replicate the distribution characteristics of the test scores with a larger sample of participants.

Method

Study 1: Development and Construct Validation

The construction of the MiniPONS was part of a validation study for a new Multimodal Emotion Recognition Test (MERT; Bänziger et al. 2009). The full version of the PONS was administered to 74 participants in this study. Based on these results, a number of items of medium difficulty were chosen to represent six of the original 11 channels of the PONS. The resulting short form was tested 6 weeks later with the same group of participants. A number of other nonverbal tests were administered to examine construct validity.

Selection of Items with Medium Difficulty from Selected Channels of the Full PONS

In the spirit of Hall's (2001) conclusions, it was decided to include a number of single and combined channels from the original test, in particular the three main channels, body only, face only, and two versions of voice only (content-filtered speech, CF, and randomized-splice speech, RS). In addition, two combined face and voice channels (face + CF and face + RS) were chosen, considering that these would cover the major nonverbal cues as represented in the literature and constitute an ecologically valid combination of cues. In consequence, potential items for the short form were chosen only out of those channels of the full PONS. An attempt was made to choose a similar number of items for each of these channels in such a way that the total number of items would allow a limited test duration of about 15 min. Sixteen items were chosen for each of the two single-channel conditions of body only and face only, and eight items each for the single-channel speech conditions (RS and CF) and the two combined-channel conditions (face + RS and face + CF), with the same number of items (two or four, respectively) for each of the quadrants of the 2 × 2 (valence by dominance) design of the PONS. This procedure yielded a total of 64 items.

The selection criteria, based on the results of the full PONS administration in the sample described above, were item difficulty (percentage of the total sample that gave the correct answer to the item) and the required number of items per channel. Item difficulty was chosen as a criterion in the interest of increasing the differentiating power of the test. Test items that almost everyone gets right or wrong are relatively inefficient and add to the length of the test without maximally differentiating degrees of competence of the participants. The easiest and most difficult items were therefore eliminated for each channel in such a way as to obtain an optimal selection given the constraint of representing each subtest with one quarter of the items representing each quadrant of the design. The 64 items retained ranged from the most difficult being answered correctly by 45% of respondents, and the least difficult being answered correctly by 100% of respondents (the 25th percentile was 71%, the median was 82%, and the 75th percentile 91%).

Tests Used for Construct Validation

Diagnostic Analysis of Nonverbal Accuracy (DANVA) We used a digital version of the DANVA (DANVA2-AF and DANVA2-AP; Baum and Nowicki 1998; Nowicki and Duke

1994) with instructions and response items translated to French. These DANVA forms include 24 photographs of facial expressions and 24 audio recordings of vocal expressions. Facial expressions are portrayals of four emotion categories (anger, fear, happiness, sadness) with two intensities (weak or strong). The test includes three facial portrayals for each combination of emotion and intensity. Facial portrayals recognized by more than 80% of the respondents in a pilot study were selected for this test by its developers (Nowicki and collaborators). The facial encoders are young Americans (not professional actors), each of whom appears maximally two times in the selected portrayals. Vocal portrayals are produced by two professional actors (one male, one female) and represent the same four emotions with two intensity levels. The emotions are always portrayed using the following sentence: "I'm going out of the room now, and I'll be back later." Vocal portrayals recognized by more than 70% of the respondents in a pilot study were selected by the developers for the test. For both tests, participants select one of four categories (anger, fear, happiness, or sadness) for each portrayal.

The 24 facial portrayals were displayed first by a computer program, in a fixed order with each photograph shown for 2 s. The 24 vocal portrayals were then presented in a fixed order as well. Participants were requested to select one of four response alternatives on screen and their answers were recorded by the computer program. Accuracy scores were computed separately for facial and for vocal scores as a proportion of correct answers (i.e., answers matching the target category).

Japanese and Caucasian Facial Expressions of Emotion (JACFEE) Test The JACFEE set of facial expressions (Biehl et al. 1997) includes photos of 14 Caucasian males, 14 Caucasian females, 14 Asian males, and 14 Asian females portraying one of seven basic emotions with configurations of facial features following the encoding instructions of Ekman et al. (1983). This dataset includes seven emotions: surprise, sadness, anger, happiness, fear, disgust, and contempt. Each emotion is portrayed with identical facial muscle contractions by eight different encoders. These 56 JACFEE photos of posed facial expressions were used in an extensive intercultural study, showing high recognition rates across all cultures studied. Using the same 56 JACFEE photos Matsumoto et al. (2000) constructed a brief affect recognition test, JACBART, in which each expressive picture is displayed for a brief period of time sandwiched between the same face in a neutral state, without expression. As the format of JACBART did not lend itself to our purposes, we created a comparable test using the same 56 JACFEE photos and exposing the expression for 1/5 of a second, which is long enough to consciously see the expressive face, between the two neutral faces. The 56 sequences, neutral-expressive-neutral, were presented in a fixed order on a computer screen and were followed by the seven alternative answers listed above. The answers selected by the participants for each picture sequence were recorded by the computer.

Emotion Recognition Index (ERI) The ERI test (Scherer 2007; Scherer and Scherer 2011) uses pictures from the Ekman and Friesen Pictures of Facial Affect (PFA) series for a 30-item/5 categories (anger, fear, joy, sadness, disgust) facial recognition test (FACIAL-INDEX) and vocal emotion portrayals selected from a large corpus of portrayals produced by four German professional radio actors (see Scherer et al. 1991) for a 30-tem/5 category (sadness, fear, anger, joy, neutral) vocal recognition test (VOCAL-INDEX), both without time limitation, presented automatically on a computer screen. All answers are recorded by the computer. The test has been validated with 1,380 employees in international companies at various levels of management (see Scherer and Scherer 2011).

The Multimodal Emotion Recognition Test (MERT) This new instrument was developed by Bänziger et al. (2009) to measure emotion recognition ability as a component of emotional competence using multimodal dynamic rather than single modality static expressions as in most past instruments. It includes 10 actor-portrayed emotions (anxiety, panic fear, happiness, elation, cold anger, hot anger, sadness, despair, disgust, and contempt), which represent two variants each for five major emotion families (differing on the arousal/intensity dimension). Each emotion is instantiated by three film clips and presented in four modes: video only (facial cues), audio only (vocal cues), audio/video (integrating facial and vocal cues), and still photographs (extracted from the film clips). Film clips, sounds, and still pictures are presented on screen followed by response alternatives and the selected answers recorded by a computer.

Participants

Seventy-four participants (64 females, 10 males), with a mean age of 22 years ($SD = 4$ years), took part in this study. They were undergraduate psychology students who participated for course credit. They were also promised and given personalized feedback on their results for all nonverbal sensitivity tests (MERT, DANVA, ERI, JACFEE, and PONS). This study took place in a large computer room at the university of Geneva, Switzerland.

Procedure

The tests and questionnaires in this study were all administered on individual computers. The study took place in a computer room habitually reserved for the students to work on their exam papers or do research on the Internet. The computers were equipped with standard headphones. The sound level was equal on all computers and adjusted for the amplification of the test stimuli. Participants took the tests in groups of 10–20. Every participant was expected to return for three separate sessions. In a first session (with 72 participants, 63 females) MERT was administered and was followed by a request to fill out a series of questionnaires, which included several self-report measures of personality (see Bänziger et al. 2009).

In the second session (the following week), the participants were requested to complete the other tests of nonverbal sensitivity, in particular the full version of the PONS ($n = 68$ participants from the original pool). In addition, ERI Vocal ($n = 68$) and Facial ($n = 72$), DANVA ($n = 70$), and JACFEE ($n = 67$) were given.

About 6 weeks after this session ($n = 67$ participants from the original pool), the newly constructed short form of the PONS, the MiniPONS instrument described in this article, was administered (together with a retest of the MERT instrument).

The tests and questionnaires were run from a distant server and results were automatically uploaded to the server. There were no technical problems reported during the sessions but a few result files were not correctly uploaded on the distant server. Some participants also failed to show up at one or sometimes two out of three sessions, which resulted in some missing data points. For this study, we included all 69 participants (59 females, 10 males) from whom we got complete answers on the MiniPONS (in the third session), but because of data loss on server upload, only 65 of them (9 males) are included in the correlations we report with instruments that were filled out on separate occasions (including correlations with the full PONS).

Study 2: Replication and Reliability Check

Given that in Study 1 the same participants took both the full version of the PONS and, several weeks later, the MiniPONS short version, one cannot exclude the possibility that the prior exposure to all items of the test has led to a learning effect which may have artificially increased the overall level of accuracy. Thus, to check the stability of the distribution measures of the test scores, we ran a second, study using the web administrated version of the MiniPONS, to examine the reliability of the test score distribution in a second, larger and more balanced, sample.

Participants

One hundred-and-thirty-six participants (110 females, 26 males), with a mean age of 24 years ($SD = 6$ years), took part in this study. They consisted of both students from different faculties and non-students who were paid for their participation as part of two ongoing studies collecting data for other purposes. None of them had taken the full PONS or the MiniPONS before.

Procedure

The web version of the MiniPONS was administered on individual computers, either as part of a group experiment in the computer lab of the university or as part of a participant panel in which members participated in web-based studies under controlled conditions in their homes or offices.

Results

Study 1: Development and Construct Validation

Reliability

Internal consistency While it is generally assumed that higher internal consistency contributes to higher validity, it is often overlooked that internal consistency is made up of two factors: (a) n, the number of items in the scale, and (b) r_{xx}, the average correlation among items. When each of the individual items has some, even very modest, correlation with the criterion variable, the validity of the scale increases as n increases, but validity decreases as the item-to-item correlation increases. In short, one would like to use a scale in which (a) each item has a bit of validity, (b) there are more items of the same type, and (c) individual items show very *low* correlations with other individual items (see Table 4:18 in Rosenthal and Rosnow 2008, p. 121). In this respect, the internal consistencies as measured by Intraclass Correlation Coefficients (ICC) were satisfactory both for the full

Table 1 Study 1: Accuracy scores for the individual channel scores and the total score; test–retest correlations

Modality	Channel	Number of items	*M*	*SD*	Minimum	Maximum	Retest reliability (r)
Audio	RS	8	0.74	0.18	0.12	1.00	.38**
	CF	8	0.78	0.15	0.25	1.00	.31*
Video	Body	16	0.77	0.12	0.50	1.00	.09
	Face	16	0.82	0.10	0.56	1.00	.28*
Both	Face + RS	8	0.80	0.14	0.38	1.00	.25*
	Face + CF	8	0.90	0.10	0.62	1.00	.29*
Total	Total	64	0.80	0.07	0.52	0.95	.64**

RS randomized-spliced speech, *CF* content-filtered speech, *M* mean proportion accuracy. $N = 69$

* $p < .05$; ** $p < .01$

PONS (195 items[1] and 70 participants) and for the MiniPONS (61 items[2] and 65 participants). For the full PONS, the Single Items ICC was .015 and the Combined Items ICC was .750, and for the MiniPONS the Single Items ICC was .021 and the Combined Items ICC was .566. Single Items ICCs are an approximation of the average correlation between items (which was .018 for full PONS and .029 for MiniPONS). Combined Items ICCs are analogous to Cronbach's Alpha, i.e., the reliability of the entire test or subtest (Rosenthal and Rosnow 2008). The results indicate that while the items are not strongly inter-correlated overall (small values for Single Items ICC), they are still related to a common underlying latent variable (relatively high values for Combined Items ICC). The larger Combined Items ICC for full PONS than for MiniPONS is due to the larger number of items (195 "valid" items for the full PONS, against 61 "valid" items for MiniPONS).

Test–retest The item selection was based on the administration of the full PONS 6 weeks before administering the MiniPONS. This allows estimating the test–retest reliability for the 64 MiniPONS items. The scores of the 65 participants who completed both tests were correlated ($r = .70$) when correlating Full-PONS scores with MiniPONS scores. The test–retest correlation was comparable when computing the two scores on the basis of the 64 items in the full PONS that were selected for inclusion in the MiniPONS ($r = .64$). The last column of Table 1 shows the test–retest correlations for the single-channel scores computed on the basis of the selected items only.

Accuracy for the MiniPONS Channels and Total

Table 1 shows the descriptive statistics for the total score of the new short form and its channels. The data suggest that the presence of facial information tends to boost accuracy. While combined channels were generally decoded more accurately than single vocal channels or body, this was not the case for face only, except in combination with content-filtered (CF) speech, which seems to add additional accuracy to the face stimuli. Figure 1

[1] PONS features 220 items, but 25 items had zero variance (all participants selected the correct answer) and could not be used in the analysis.

[2] MiniPONS features 64 items, but 3 items had zero variance (all participants selected the correct answer) and could not be used in the analysis.

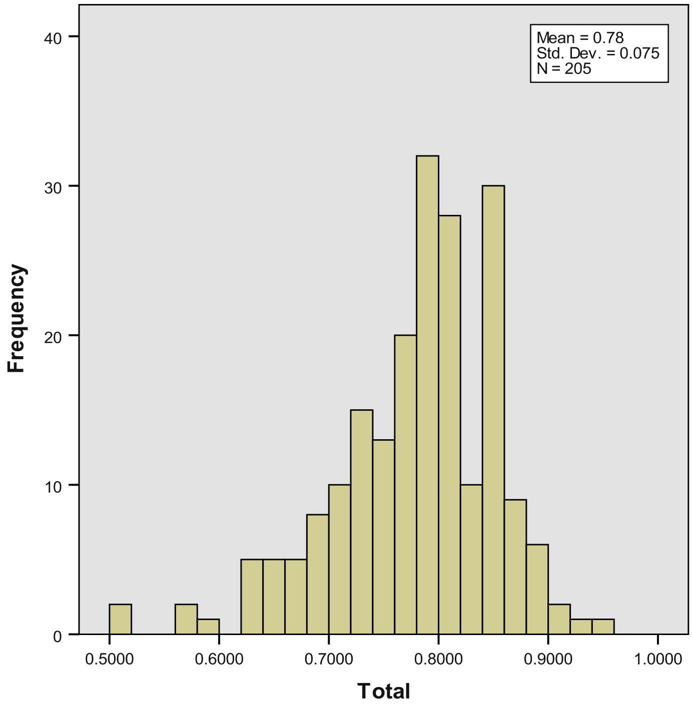

Fig. 1 Distribution of total scores of the MiniPONS in combined Study 1 and 2

presents the distribution for the total accuracy scores across all subtests, which approximates a normal curve.

Correlation of the MiniPONS with the Full PONS

Table 2 shows the correlations of the channels and the total score of the MiniPONS with the channels and the total score of the full PONS. The correlations for the equivalent channels and total scores are highlighted in bold. As can be seen, these do not in all cases correlate significantly or with a higher *r* than correlations with nonequivalent channels. This pattern can be explained by two major factors: (a) for four of the six MiniPONS channels only eight items were used, reducing the stability of the scores, (b) there are multiple overlaps of nonverbal cues in different subsets which makes a direct comparison of the channel scores difficult. Thus, face is involved in three of the channels and it is not surprising that the face-only condition correlates very highly with the combined channel scores. Also, the two vocal channels focus on different aspects of speech (RS on voice quality, CF on dynamic speech cues) but they share some cues such as pitch. However, the pattern of correlations suggests that channels should not be used on their own but rather as an integral part of the whole test, assessing different types of recognition competences. As mentioned in the Introduction, the strength of this multimodal short form of the PONS is that it includes several channels and channel combinations at the same time. As the last column of Table 2 shows, the correlations of the MiniPONS channel scores with the total score of the full PONS are quite sizeable and highly significant, suggesting a reasonable amount of shared variance. Most importantly, the correlation between the total scores of

Table 2 Study 1: Correlations between channels of the MiniPONS and Full PONS

MiniPONS	Full PONS: Channels from which MP items were drawn						Full PONS: Channels from which no MP items were drawn					
	P RS	P CF	P body	P face	P fa + RS	P fa + CF	P bo + RS	P bo + CF	P fig	P fig + RS	P fig + CF	P total
MP RS	**.22***	.19	.27*	.28*	.05	−.09	.09	.26*	.20	.37**	.29**	.38**
MP CF	.23*	**.20**	.31**	.21*	.20	.06	.27*	.15	.14	.23*	.15	.38**
MP body	.22*	.11	**.03**	.19	.07	.08	.13	.26*	.25*	.19	.25*	.31**
MP face	.11	.01	.25*	**.26***	.40**	.27*	.18	.31**	.17	.27*	.36**	.44**
MP face + RS	.07	.00	.33**	.20	**.29***	.31**	.39**	.18	.30**	.36**	.32**	.46**
MP face + CF	.36**	.15	.35**	−.01	.24*	**.27***	.30**	.07	.19	.33**	.22*	.44**
MP total	.34**	.19	.41**	.36**	.35**	.24*	.37**	.39**	.37**	.50**	.48**	**.70****

P PONS full version, *MP* MiniPONS, *RS* randomized-spliced speech, *CF* content-filtered speech, *bo* body, *fa* face, *fig* total figure (face + body)

* $p < .05$, one-tailed; ** $p < .01$, one-tailed; $N = 65$

Table 3 Study 1: Correlations between channels of the MiniPONS and of the corresponding channels of the construct validation tests

	PONS face	MP face	DANVA face	MERT photo	MERT video	JACFEE	ERI facial
(a) Video presentation of facial expression							
MP face	.26*						
DANVA face	–.14	.12					
MERT photo	.04	.20	.22*				
MERT video	.18	.23*	.11	.54**			
JACFEE	.11	.31**	.28*	.33**	.27*		
ERI facial	–.14	.23*	.15	.32**	.06	.29*	
Mean *r* for all seven tests (columns)	.05	.22	.12	.28	.23	.26	.15
Median *r* for all seven tests (columns)	.08	.23	.14	.27	.20	.28	.19

	PONS voice	MP CF	MP RS	DANVA voice	MERT voice	ERI vocal
(b) Audio presentation of vocal expression						
MP CF	.26*					
MP RS	.24*	.31**				
DANVA voice	.24*	.05	.17			
MERT voice	.25*	.28*	.13	.32**		
ERI vocal	.03	.09	.03	.01	.27*	
Mean *r* for all six tests (columns)	.20	.20	.18	.16	.25	.09
Median *r* for all six tests (columns)	.24	.26	.17	.17	.27	.03

	PONS FaVo	MP face + CF	MP face + RS	MERT AV
(c) Audio-visual presentation of concurrent facial and vocal expression				
MP face + CF	.33**			
MP face + RS	.39**	.40**		
MERT AV	.34**	.19	.21*	
Mean *r* for all four tests (columns)	.35	.31	.33	.25
Median *r* for all four tests (columns)	.34	.33	.39	.21

MP MiniPONS, *RS* randomized-spliced speech, *CF* content-filtered speech, *Vo* voice, *Fa* face, *AV* audio-visual. Further information about the validation tests is provided in the Method section. Further details about all of these tests, particularly on the number of items in the respective subtests and the way in which total scores are computed, can be found in Bänziger et al. (2009)

* $p < .05$, one-tailed; ** $p < .01$, one-tailed

the short and the long form is highly satisfactory ($r = .70$),[3] demonstrating that the short form shares a major portion of the variance with the full form. This is in the ballpark for group administered paper-and-pencil IQ test–retest reliability (Thorndike 1933). In conclusion, the overall validity of the MiniPONS as a short form of the full PONS is very substantial.

[3] A correction for attenuation was computed for this correlation using the reliability estimates provided in the section *Internal Consistency*. Based on those estimates (.750 for PONS and .566 for MiniPONS) the corrected (disattenuated) correlation would be larger than 1 (1.07).

Table 4 Study 1: Correlations of total scores on MiniPONS and full PONS with the four validation tests

PONS version	DANVA	ERI	JACFEE	MERT
MiniPONS total	.15	.24*	.48**	.32**
PONS total	.14	.09	.59**	.51**

See text for explanation of acronyms

* $p < .05$; ** $p < .01$

Correlations with Construct Validity Tests

Table 3 shows the correlations of the MiniPONS channels and the corresponding channels of the other emotion recognition tests used for construct validation. Not all of the channels that could reasonably be expected to correlate highly show significant correlations. However, it should be noted that most of the other tests, except the MERT, use only static pictures (photographs) for the facial channel. Furthermore, they are all based on a few basic emotions (except MERT) whereas PONS is a more general nonverbal sensitivity test, which includes attitudes and social relationships, thus differing strongly with respect to test design and construction and the response options used.

Given these major differences, there is an impressive degree of correlation between the MiniPONS channels and the corresponding channels in the validation tests, suggesting that the short form of the PONS can be expected to have high validity. Table 3a shows the correlations for video presentations of facial expressions. The MiniPONS Face subscore shows lower correlations than the JACFEE or MERT-Photo but higher correlations than the ERI and DANVA Face tests. Importantly, the MiniPONS Face score clearly outperforms the full PONS Face score in this respect. In the vocal channel (see Table 3b), the MERT Voice score clearly captures the most common variance with the other tests, but the MiniPONS CF score correlates about as highly with the MERT Voice score as do the DANVA and full PONS Voice scores (the latter representing a combination of CF and RS in this table, a choice made for the sake of economy). For audio–video face plus voice presentations (see Table 3c), only the two PONS versions and the MERT AV score qualify as the other tests do not contain concurrent vocal and facial expressions. While lower than the full PONS Face and Voice score, the two MiniPONS scores (CF and RS) still correlate

Table 5 Study 2: Accuracy scores for the individual channel scores and the total score of the MiniPONS

Modality	Channel	Number of items	Females ($n = 110$)				Males ($n = 26$)				Total ($N = 136$)			
			M	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max
Audio	RS	8	0.80	0.16	0.13	1.00	0.75	0.20	0.25	1.00	0.79	0.16	0.13	1.00
	CF	8	0.78	0.15	0.25	1.00	0.73	0.16	0.25	1.00	0.77	0.15	0.25	1.00
Video	Body	16	0.75	0.09	0.50	1.00	0.68	0.11	0.50	0.94	0.74	0.10	0.50	1.00
	Face	16	0.74	0.10	0.50	0.94	0.72	0.11	0.50	1.00	0.74	0.10	0.50	1.00
Both	Face + RS	8	0.76	0.16	0.29	1.00	0.74	0.19	0.29	1.00	0.75	0.17	0.29	1.00
	Face + CF	8	0.89	0.12	0.50	1.00	0.82	0.18	0.38	1.00	0.88	0.13	0.38	1.00
Total	Total	64	0.78	0.06	0.57	0.89	0.73	0.10	0.51	0.91	0.77	0.07	0.51	0.91

RS randomized-spliced speech, *CF* content-filtered speech, *M* mean accuracy, *SD* standard deviation, *Min* minimum (lowest individual score), *Max* maximum (highest individual score) reasonably well with the MERT (especially given the reduced number of cues in comparison to the combined score for the full PONS).

The correlations of the channel scores with the validation tests are reported here for the sake of completeness. However, as mentioned above, the channel scores should not be interpreted for diagnostic purposes. Given the small number of items and the cue overlap between different items, it may not be appropriate to interpret scores at the level of individual channels.

In conclusion, given the respectable correlation of $r = .70$ with the full version, the MiniPONS short form seems to capture an essential portion of the variance. With respect to construct validity, the results reported in Table 3 and discussed above are also quite respectable. However, given the small number of items per channel in the MiniPONS, it may be safer to evaluate the construct validity on the basis of the total score. Table 4 shows this central piece of evidence—the correlations between the short and long form scores and the tests used for construct validation. While the MiniPONS had lower correlations with the JACFEE and the MERT than did the full PONS, these remained highly significant. At the same time, while the long form did not correlate significantly with the ERI, the short form did. The DANVA did not correlate significantly with either PONS version, which parallels similarly low correlations with the MERT (see Bänziger et al. 2009).

Study 2: Replication and Reliability Check

Accuracy for the Channel Scores and the Total Score

Table 5 shows the descriptive statistics for the MiniPONS scores in Study 2. While the total score is lower than in Study 1, the difference is slight, demonstrating a high degree of stability of the score and ruling out strong learning effects in Study 1. Furthermore, the accuracy profile over channels is very similar between the two studies, with the combination of facial information with content-filtered (CF) speech producing particularly high accuracy. The differences between the two studies are likely to be due to sampling factors, including the larger proportion of males in Study 2. A *t*-test for gender differences shows a significant effect favoring females in Study 2 ($t(134) = -3.06$, $p = .003$, with a difference of .05 between the respective means (effect size $r = .25$; $d = .52$). The individual channel scores and the total scores for the genders are described in Table 5, showing that the profiles over channels are also very similar across gender. In consequence, it seems advisable to combine the samples from both studies to have a larger basis of cases. Figure 1 shows the overall distribution of the total score (for the combined samples) in the form of a histogram, which approximates a normal curve.

Discussion

The development of a short form of the PONS as described in this article has a number of limitations. One concern is that measures of basic emotion recognition are not ideally suited to establish construct validity for a more general nonverbal sensitivity test based on social situations of different types of activity. However, as there is no readily available test that could be directly compared to the PONS, tests of emotion recognition seemed the best alternative option. A further limitation is the relatively small size of our samples and the uneven gender ratio of the participants, which implies that the results we describe can be generalized with more confidence to women than to men. The current work ought to be complemented by further assessments with larger samples—including more males—in particular with respect to establishing the predictive validity of the test and comparing it to that of other tests.

In conclusion, the full PONS has been a highly successful test of nonverbal sensitivity with clearly demonstrated predictive validity. One of its drawbacks has been its length and difficult administration. This article demonstrates that a 64-item short form (MiniPONS) yields a rather impressive correlation with both the full PONS test and several other tests of emotion recognition. The reduction of the length of the test (64-items against 220-items for the full version) necessarily results in lower reliability of the scores. The well known trade-off between length/speed and reliability also applies to the comparison between the PONS and the MiniPONS.

The MiniPONS test described here can be administered on the web or in the laboratory and is freely available to researchers or to all those who want to obtain a rapid screening of their nonverbal sensitivity.[4]

The nonverbal sensitivity to affective communication by others is a major part of emotional intelligence (or rather *competence*, see Scherer 2007). A central component of such competence is to correctly infer the complex affective states that individuals experience and communicate in specific social situations which include fairly standardized communicative actions like requesting, commanding, scolding, pleading, etc. This is what the PONS assesses and a multitude of studies have consistently demonstrated moderate but very stable and consistent predictive effects. Given the high correlation with the full test and the convergent construct validity reported above, the MiniPONS is likely to render the use of this promising instrument more accessible for researchers in many areas. However, endorsing Bollen and Lennox's (1991) distinction between *effect* indicator and *causal or constituent* indicator models, we wish to assert our firm conviction that emotional competence as a latent trait needs to be modeled as an emergent constituent indicator model rather than an invariant trait that will have a consistent effect on many interchangeable indicators. The MiniPONS measures one such constituent indicator, nonverbal sensitivity to complex affective communication in social situations. Emotion recognition tests, which ask participants to identify modal emotions (Scherer 1994) in a completely context-free fashion, measure a different constituent. As the results reported in this article show, both approaches share variance, but the overlap rarely exceeds 20%. Even among the different emotion recognition tests, the shared variance rarely exceeds this level. Most likely even in such relatively straightforward recognition tasks, there may be sub-constituents of the emergent latent trait. In fact, Bänziger et al. (2009) reported for the MERT that a factor analysis of the data suggests two separate abilities, visual and auditory recognition (largely independent of personality dispositions). In consequence, in order to assess an emergent latent trait of emotional competence, it will be essential to develop a battery of tests that allows aggregating different, albeit overlapping, constituents.

Acknowledgments The research reported here was supported by the Swiss National Science Foundation, and the Swiss Center for Affective Sciences. The authors gratefully acknowledge important contributions by Didier Grandjean, Fridtjof Nussbeck, and Olivier Rosset.

[4] The web version of the MiniPONS can be freely accessed and individuals completing the test will receive feedback on their performance. Go to http://www.affective-sciences.org/webexperimentation and choose MiniPONS; English, French, and German versions are available. A version of the MiniPONS is available for researchers to test groups of participants. Researchers interested in using the MiniPONS for group administration, any of the PONS short forms or the full PONS should contact Dr. Judith Hall (j.hall@neu.edu).

References

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*, 947–961.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518–529.

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*, 691–704.

Baum, K., & Nowicki, S., Jr. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior, 22*, 89–109.

Bernieri, F. J. (1991). Interpersonal sensitivity in teaching interactions. *Personality and Social Psychological Bulletin, 17*, 98–103.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3–21.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

DiMatteo, M. R., Friedman, H. S., & Taranta, A. (1979). Sensitivity to bodily nonverbal communication as a factor in practitioner-patient rapport. *Journal of Nonverbal Behavior, 4*, 18–26.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*(4616), 1208–1210.

Funder, D. C., & Harris, M. J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality, 54*, 528–549.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*, 845–857.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: The Johns Hopkins University Press.

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. 143–160). Mahwah, NJ: Erlbaum.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*, 149–180.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology, 77*, 350–359.

Hodgins, H. S., & Koestner, R. (1993). The origins of nonverbal sensitivity. *Personality and Social Psychology Bulletin, 19*, 466–473.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). Oxford, UK: Oxford University Press.

Kurkul, W. W. (2007). Nonverbal communication in one-to-one music performance instruction. *Psychology of Music, 35*, 327–362.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., et al. (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*, 179–209.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy. *Journal of Nonverbal Behavior, 18*, 9–35.

Robbins, J. M., Kirmayer, L. J., Cathébras, P., Yaffe, M. J., & Dworkind, M. (1994). Physician characteristics and the recognition of depression and anxiety in primary care. *Medical Care, 32*, 795–812.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: The Johns Hopkins University Press.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). New York: McGraw-Hill.

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior, 28*, 267–286.

Scherer, K. R. (1994). Toward a concept of "modal emotions". In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 25–31). New York/Oxford: Oxford University Press.

Scherer, K. R. (2007). Component models of emotion can inform the quest for emotional competence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 101–126). New York: Oxford University Press.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123–148.

Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. *Journal of Psycholinguistic Research, 14*, 409–425.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index (ERI). *Journal of Nonverbal Behavior*.

Thorndike, R. L. (1933). The effect of the interval between test and re-test on the constancy of the IQ. *Journal of Educational Psychology, 24*, 543–549.