Children Australia

www.cambridge.org/cha

Development of facial emotion recognition and empathy test (FERET) for primary school children

Kerem Coskun

Department of Primary Teaching, Faculty of Education, Artvin Coruh University, Artvin, Turkey

Article

Cite this article: Coskun K (2019) Development of facial emotion recognition and empathy test (FERET) for primary school children. *Children Australia* **44**: 23–31. https://doi.org/10.1017/cha.2018.51

Received: 26 February 2018
Revised: 25 October 2018
Accepted: 5 December 2018
First published online: 18 February 2019

Keywords:
Facial emotion recognition; emotional skills; empathy; test development; primary school children

Author for correspondence: Kerem Coskun, Email: keremcoskun@artvin.edu.tr

Abstract

This article reports on the development of the Facial Emotion Recognition and Empathy Test (FERET) as a reliable and valid tool for assessing facial emotion recognition and empathy skills in primary school-aged children. Pictures of human faces developed by the researcher were used as response options for the children. The range of response options and their associated scores were constructed based on the Two Factorial Consensual Mood Structure, which indicates dimensions of emotions. Four hundred and twenty-two primary school children participated in the research. The children were asked to recognise emotions and display the appropriate empathetic response. Data were analysed through item analysis, exploratory factor analysis (EFA) and Item Response Theory (IRT). As a result of item analysis, FERET has been designed to include 6 items related to facial expressions of key emotions and has been found to have an internal consistency coefficient of .82 and be unidimensional. Results from the IRT indicated that all the items in FERET discriminate better responders from poor responders. It was concluded that FERET can produce reliable and valid results in measuring facial emotion recognition and empathy skills among primary school children.

Introduction

Emotions are regarded as inseparable from, and important aspects of, human beings, and humans have various basic, innate emotions that are shared across cultures and play vital functions in adaptation (Crowley, 2014). Emotions can be expressed through the body and each has distinctive bodily manifestations. Emotions stem from mental perceptions of a stimulus or object, which excites mental affection (Cabanac, 2002), and trigger changes in the body in order to respond to environmental demands. Therefore, they have a crucial function in social adjustments and it is important for individuals to be able to accurately recognise emotions. Most people learn emotion recognition skills by naming emotions and developing an emotional lexicon based on facial expressions (Matthews, 2006). Furthermore, people learn to label and recognise emotions in themselves and others.

Across communities, emotions are usually expressed using the same facial musculature (Ekman & Friesen, 1971, 1988). Facial emotion recognition is therefore an important component of communication and successful human interactions, which relies on individuals' abilities to decipher facial expressions (Pollak & Sinha, 2002). Facial emotion recognition can be described as the skill of decoding emotions in others by examining human faces (Rubin, Munz, & Bommer, 2005). Facial emotion recognition depends on dynamic modulation of face musculature in others to predict which emotion is being presented, and this process of prediction enables subconscious mapping of the physical features of the face; hence, it is possible to accurately observe what others feel. Facial emotion recognition, in turn, allows people to generate appropriate responses to others. Consequently, accurate facial emotion recognition is an important part of adaptive social functioning (Pollak & Sinha, 2002).

However, facial emotion recognition is not enough for individuals to be able to adjust to the social environment. Empathy is another skill that enables adaption to social settings. Empathy is a phenomenon that has been conceptualised in a number of ways by different theoretical models. Mehrabian and Epstein (1972) described empathy as an emotional response based on perception of emotions in others. Facial emotion recognition can therefore be seen as an essential component of empathy. Davis (1983) similarly defined empathy as the reactions based on observed experiences of others, while Dymond (1949), Feshbach and Roe (1968), Milgram (1960) and Strayer (1987) described empathy as the ability to predict others' behaviours and emotional conditions based on cues of social situation and personality.

These descriptions imply that empathy has a cognitive aspect rather than being solely an affective concept, and certainly perception of emotions in others is a cognition-laden skill (De Vignemont & Singer, 2006; Dymond, 1949, 1950; Stueber, 2010; Zaki, Bolger, & Ochsner, 2009). Furthermore, empathy operates based on social referencing, which is defined

CAMBRIDGE UNIVERSITY PRESS

as the use of other peoples' emotional expressions to respond to uncertain and vague situations. In other words, social referencing is a prerequisite for showing empathy. In addition, empathy is such an emotional concept that it refers to the ability to experience and share emotions with others (Mehrabian & Epstein, 1972). The ability to relate to others' emotional states is crucial in terms of emotional connectedness, social functioning and regulation of social interactions. In affective terms, empathy is an emotional state that is triggered and affected by another's emotional state and arousal (De Waal, 2008). Given that the cognitive aspect of empathy emphasises perception of emotion in others, the affective aspect of empathy addresses empathy as a skill to experience and share emotions with others. Consequently, empathy can be viewed as a multifaceted concept.

Within the first 2 years of life, children can display empathetic behaviours in response to another's emotional state. Although young children have limited cognitive capabilities, they can experience empathy because they can be aroused through mimicry, conditioning and association. As children mature, their capacity to empathise is increasingly shaped by their social–cognitive development rather than mimicry, conditioning and association (Decety & Jackson, 2004). In late childhood, for example, children develop a sense of self and become able to establish boundaries between self and others. As a result of developmental advancements, children in late childhood are able to empathise through the integration of cognitive and emotional skills, instead of relying on the stimulus-driven experience of early childhood. With this in mind, empathy discloses a lot about child development (De Waal, 2008; Hoffman, 2007). The integration of cognitive processes and affective skills means that empathy can be viewed as an indicator of neurodevelopment (Decety, 2010). Measuring the empathetic skills of primary school children can therefore provide a means of monitoring cognitive and affective development in primary school children.

The development of empathy skills among students also has positive outcomes in education. Empathy is positively associated with prosocial behaviours, social competence and academic competence, whereas it is negatively correlated with aggressive and antisocial behaviour (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Carlo, Hausmann, Christiansen, & Randall, 2003; Costa & Faria, 2015; Eisenberg & Miller, 1987; Wentzel, 1993). Indeed, empathy is considered such an important indicator of social and moral development (Hoffman, 2007) as well as socio-emotional skills that it is taught to children in school curriculums from kindergarten to high school in the context of socio-emotional learning (SEL) in developed countries (Matthews, 2006; Wentzel, 1993). Determining the impact of school-based instructional programs is a necessary part of the teaching process, and a great number of instruments assessing empathetic skills have been developed in order to measure and assess the impact of these programs. However, it is very difficult to measure facial emotion recognition and empathy skills among children objectively and reliably because the experience and perception of socio-emotional skills are fundamentally subjective. In psychological testing, three methods are used for assessment and measurement: objective performance tests, self-reports and behaviour observation (Murphy & Davidshofer, 2005). The first of these, measurement of facial emotion recognition and empathy skills through self-report, has several drawbacks. Self-reporting obviously depends on self-appraisal and can be easily influenced by personal dispositions, motivations and social vulnerability – self-report assessment is thus subjective rather than objective (Goldenberg, Matheson, & Mantler, 2006). Behaviour observation, which depends on systematic observation in natural settings, can also be a very difficult process, and, like self-reporting, may be too subjective. Performance-based assessment tests, on the other hand, depend on an evaluation of responses or performances in terms of previously determined criteria, making performance-based measurements more objective. Performance-based measurements remove the possibility of the test-taker's bias in rating their facial emotion and empathy skills. It therefore enables measurement of the test-taker's performance in a more reliable and valid way (Goldenberg et al., 2006; Mayer, 2001; Mayer, Salovey, & Caruso, 2004).

Purpose of the Study

Emotion recognition and empathy skills have been conceptualised using a variety of constructs such as emotional intelligence, emotional literacy, social competency and social–emotional skills, and there are a lot of instruments that assess facial emotion recognition and empathy based on these constructs. Humphrey et al. (2011) and Wigelsworth, Humphrey, Kalambouka, and Lendrum (2010) have reviewed and identified the instruments that assess emotion recognition and empathy in the context of other constructs.

In addition to the instruments identified in Table 1, Matsumoto and Ekman (1988) have developed the Japanese and Caucasian Facial Expression of Emotion Scale (JACFEE) as an instrument to measure emotion recognition of adults through self-report. However, as indicated in Table 1, there does not appear to be an instrument that measures both emotion recognition and empathy skills among primary school children. The aim of the present study was to develop such an instrument in order to resolve this gap in the research.

Method

The present study took the form of survey research because generalisability of the research findings outside the research sample was one of the main objectives (Fraenkel, Wallen, & Hyun, 2012). The aim of the study was to develop a performance-based instrument that yields reliable and valid results in assessing facial emotion recognition and empathy skills among primary school children. Test development consisted of two phases: (1) construction of the test and (2) norming and standardising the test (Murphy & Davidshofer, 2005).

Construction of the Test

Test construction included selection of emotion categories, item content, item response and scale construction.

Selection of emotion categories

Assessment of facial emotion recognition and empathy skills through performance-based testing is very difficult because it requires objective criteria to determine whether responses are correct or incorrect. A highly structured item type was therefore chosen to assess facial emotion recognition and empathy skills among primary school children. The structured response method is useful for designing tests that can assess a broad and rich range of human abilities and skills (Murphy & Davidshofer, 2005). As emotion recognition and empathy skills were to be assessed, emotions which are to be recognised and their empathetic responses were depicted on human faces due to the nature of the performance-based assessment. The Two Factor Consensual

Table 1. Review of instruments for use with children and adolescents

Measure	Acronym	Way of measurement	Age range (years)	Domain
Assessment of Children's Emotional Skills (Schultz, Izard, & Bear, 2004)	ACES	Maximal	4–8	Emotional
Child Assertive Behaviour Scale (Michelson & Wood, 1982)	CABS	Maximal	2.5–6.5	Emotional
Child Rating Scale (Hightower et al., 1987)	CRS	Self-report	5–13	Social–emotional
Children's Self-Report Social Skills Scale (Danielson & Phelps, 2003)	CS4	Self-Report	9–12	Social
Diagnostic Assessment of Non-Verbal Accuracy (Nowicki & Duke, 1989).	DANVA	Maximal	4 and over	Emotional
Differential Emotions Scale (Izard, Dougherty, Bloxom, & Kotsch, 1974)	DES	Self-report	7 and over	Emotional
Emotional Awareness Questionnaire (Rieffe et al., 2007)	EAQ	Self-Report	9–16	Emotional
Emotional Literacy Assessment and Intervention (Faupel, 2003)	ELAI	Self-Report	7–16	Emotional
Emotional Quotient Inventory: Youth Version (Bar-On & Parker, 2008)	EQi: YV	Self-report	7–18	Social–emotional
Emotion Regulation Checklist (Shields & Cicchetti, 1997)	ERC	Self-report	6–12	Emotional
Matson Evaluation of Social Skills (Matson, Rotatori, & Helsel, 1983).	MESSY	Self-report	4–18	Emotional
Pre-school and Kindergarten Behaviour Scales (Merrell, 1996)	PKBS	Self-report	3–6	Social
Prosocial Tendencies Measure-Revised (Carlo et al., 2003)	PTMR	Self-report	11–18	Social–emotional
Social Competence and Behaviour Evaluation Scale (LaFreneire & Dumas, 1996)	SCBE	Self-report	3–6	Social
Social Development Scale (Ainley, 2006)	SDS	Self-report	11–16	Social
Social Skills Improvement System (Gresham & Elliot, 1990)	SSIS	Self-report	3–18	Social–emotional
Southampton Test of Empathy for Preschoolers (Howe, Cate, Brown, & Hadwin, 2008)	STEP	Maximal	3–5	Social–emotional
Teenage Inventory of Social Skills (Inderbitzen & Foster, 1992)	TISS	Self-report	12–18	Social
Trait Emotional Intelligence Questionnaire Adolescent Version (Petrides, Sngareau, Furnham, & Frederickson, 2006)	TEIQue	Self-report	8–16	Social–emotional
Trait Meta Mood Scale for Children (Rockhill & Greener, 1999)	TMMS-C	Self-report	8–13	Emotional
Tromso Social Intelligence Scale (Gini, 2006)	TSIS	Self-report	13–18	Social

Mood Structure developed by Watson and Tellegen (1985) was used as the theoretical system to construct items and their responses.

The Two-Factor Consensual Mood Structure sorts emotions into quadrants. Emotions in the same quadrant are highly positively correlated. Emotions 90° apart are designated as unrelated, while emotions positioned opposite to each other across the circle are negatively correlated. Key emotions were chosen from the Two Factorial Consensual Mood Structure and depicted as human faces, with their positions in relation to other emotions in the structure used to inform the scaling of emotional responses as either highly associated, moderately associated and negatively associated or unrelated.

Item content, item response alternatives and scale construction

The design of this test assumes emotion recognition to be a precursor of empathy, and therefore aimed to measure the accurate recognition of emotion in order to test empathy skills. Fifteen target emotions to be recognised and 45 emotions as correspondent choices were selected from the Two Factorial Consensual Mood Structure. This was done so that responses could be chosen from a limited set of multiple choice answers, making item scoring and evaluation of results very easy (Murphy & Davidshofer, 2005).

The resulting Facial Emotion Recognition and Empathy Test (FERET) was designed to test whether primary school children could recognise emotion in others and respond facially and appropriately by choosing one facial expression among several options. Fifteen facial emotions to be recognised were determined from The Consensual Mood Structure. The responses were chosen from the Two-Factor Consensual Mood Structure as highly positively correlated, moderately correlated and unrelated or negatively correlated. Facial expression of emotions and their responses were depicted after all were clearly defined and determined from the Two-Factor Consensual Mood Structure. Highly positively correlated responses were given 3 points, moderately correlated responses were coded as 2 points and unrelated or negatively correlated responses were given 1 point. Students with better empathy skills were expected to choose highly correlated emotions among the three correspondent emotions.

In the draft of the form, it was thought that the children should be asked to recognise emotions in the first column and chose one emotion among three options depicted through human faces. After the items were constructed, a draft of the test with 15 items was finalised.

Three experts, from the fields of emotional intelligence, fine arts and child development, assessed the first draft of the test. In response to their feedback, three named emotions (peppy, sluggish and grouchy) were discarded from the test to improve content validity. Finally, the design of the FERET was revealed. Facial emotions and their corresponding response options are indicated in Table 2.

Table 2. Facial emotions and their response options

Number of item	Emotions to be recognized	Highly positively correlated	Moderately correlated	Unrelated or negatively correlated
1	Happy	Contend	Excited	Sad
2	Kindly	Pleased	Astonished	Quiescent
3	Aroused	Astonished	Elated	Still
4	Sad	Unhappy	Drowsy	Content
5	Relaxed	Placid	Quiescent	Jittery
6	Astonished	Surprised	Scornful	Sluggish
7	Grouchy	Blue	Nervous	Satisfied
8	Unhappy	Sorry	Dull	Still
9	Elated	Enthusiastic	Surprised	Drowsy
10	Excited	Elated	Happy	Serious
11	Lonely	Sad	Fearful	Surprised
12	Fearful	Nervous	Hostile	Calm

Norming and standardising

Norming and standardising the test consisted of three steps: norming and defining the target population, the selection of the sample and recruitment of the participants and standardisation. Since the normative group or target population is determined by the intended use of the test (Murphy & Davidshofer, 2005), the target population was defined as primary school children whose ages range from 7 to 10 years. A total of 422 primary school children aged between 7 and 10 years were recruited through cluster sampling.

Official and ethical permissions were sought from the local education authority in Artvin, Turkey, to conduct the test with the children. Schools were visited after official and ethical permission had been approved. Teachers were instructed on the purpose of the research and what the children were being asked to do and only those classrooms whose teachers agreed took part in the study. The purpose of the research was explained to the participating children by the researcher and it was emphasised that they did not have to participate in the research. Children voluntarily participated in the study with their teachers' consent. No children refused to participate in the research. In order to comply with ethical considerations, the participating children were asked not to disclose any identifying information except their age and gender. Of the children who took part, 111 were 7 years old, 102 were 8 years old, 97 were 9 years old and 112 were 10 years old. The research sample consisted of 197 female and 225 male children.

Colour printouts of the test were given to the participating children. Each of 12 items was responded to in three steps. First, they were asked to look and recognise the emotion in the first column. Secondly, they were requested to imagine what they would feel if a class friend looked like the pictured emotion. And thirdly, they were told to pick one of the emotional responses among three options in the other columns. As a result, they were asked to respond to a total of 12 target emotions in order. Moreover, they were asked not to influence each other in order to minimise extraneous variables which might impact on their test performance. Testing conditions and procedures were kept as consistent as possible for all of the classrooms in the schools which were visited (Murphy & Davidshofer, 2005). Completion of the test took about 10 minutes. All of the testing procedures were conducted in the afternoon because courses in the morning are allocated to academic learning outcomes in Turkey. As a result, testing procedures and conditions were therefore standardised and remained the same.

Results

Data analysis

Psychological tests must reveal latent trait variance. There are three main strategies to disclose latent trait variance which are related to non-test manifestations, internal consistency based on inter-item correlations as structural analysis and item response theory (IRT) focusing on latent traits (Clark & Watson, 1995; Loevinger, 1957). Structural analysis and IRT were employed to construct the FERET. In other words, structural analysis was carried out through classic testing tradition and IRT, one of the modern testing approaches. Structural analysis was conducted through corrected item-total correlation, internal consistency and unidimensionality.

Structural analysis

Structural analysis was used to identify which items should be included in the test. Structural analysis also enables revealing latent structure of the instruments. This stage was salient because the FERET measures facial emotion recognition and empathy skills through theoretical-based measurement (Clark & Watson, 1995). Therefore, corrected item-total correlation was employed to reveal internal consistency and unidimensionality.

Before the reliability analysis was computed, item analysis was carried out with corrected item-total correlation in order to identify items which did not correlate overall to the test or usefully assess different dispositions and features (Field, 2009; Nunnally & Bernstein, 1994). Corrected item-total correlation helps to reveal to what extent items and total test scores measure the same trait (Murphy & Davidshofer, 2005). Item analysis results are displayed in Table 3.

Item-total correlation has a direct impact on internal consistency and precision of the test (Nunnally & Bernstein, 1994). Therefore, items whose correlation with the total score is below .30 were discarded, whereas items whose correlation with the total score is higher than .30 were included in the test analysis. Item analysis indicated that Item 1, Item 2, Item 3, Item 7 and Item 9 should be discarded from the rest of the analysis because their item-total correlation coefficients were less than .30. Item 5 was discarded because it reduced the reliability coefficient. Item 4, Item 6, Item 8, Item 10, Item 11 and Item 12 were included in the analysis because their corrected item-total correlation was found to be over .30.

Internal consistency is a reliability method

Internal consistency indicates the overall degree to which items comprising a test are intercorrelated (Briggs & Cheek, 1986). Internal consistency indicates a reliability assessment of consistency among items (Cronbach, 1951; Murphy & Davidshofer, 2005; Nunnally & Bernstein, 1994). The alpha coefficient represents reliability based on internal consistency. Internal consistency was conducted, and the internal consistency coefficient was found to be .82. It was concluded the FERET has good internal consistency.

Table 3. Result of the item analysis

Item	Scale mean if item deleted	Value of corrected-item correlation
Happy (Item 1)	26.39	.25
Kindly (Item 2)	26.69	.28
Aroused (Item 3)	26.96	−.06
Sad (Item 4)	26.50	.68
Relaxed (Item 5)	26.09	.45
Astonished (Item 6)	26.31	.43
Grouchy (Item 7)	26.47	.01
Unhappy (Item 8)	26.51	.66
Elated (Item 9)	26.83	.04
Excited (Item 10)	26.41	.61
Lonely (Item 11)	26.69	.30
Fearful (Item 12)	26.44	.43

Table 4. Exploratory factor analysis

Item	Factor loadings	M	SD	Alpha if item deleted
Sad (Item 4)	.68	2.43	.81	.76
Astonished (Item 6)	.38	2.62	.67	.81
Unhappy (Item 8)	.62	2.42	.81	.77
Excited (Item 10)	.56	2.52	.70	.79
Lonely (Item 11)	.45	2.41	.71	.80
Fearful (Item 12)	.42	2.49	.72	.80

Eigenvalues = 3.11 total variance explained: 52% KMO = .83.
Barlett Test: $X^2 = 948.904$, $p < .001$.

Table 5. GRM item parameter estimates

Item	a	b_1	b_2
Sad (Item 4)	3.31	−0.92	−0.39
Astonished (Item 6)	1.35	−2.06	−1.09
Unhappy (Item 8)	2.67	−1.00	−0.37
Excited (Item 10)	2.02	−1.56	−0.48
Lonely (Item 11)	1.40	−1.51	0.41
Fearful (Item 12)	1.33	−1.79	−0.52

Unidimensionality is another consideration when constructing tests. Unidimensionality refers to the assessment of a single factor or construct (Clark & Watson, 1995, Cortina, 1993). Moreover, unidimensionality is an essential assumption of IRT. Exploratory Factor Analysis (EFA) was conducted to identify unidimensionality. EFA is a statistical process that allows determination of inter-correlated items, and clusters them under same construct (Field, 2009; Rummel, 1967). Before undertaking the EFA, the Kaiser–Meyer–Olkin (KMO) coefficient and Barlett Test were calculated to ensure the data were large enough to conduct the EFA. The KMO coefficient was found to be .83 and the Barlett Test was significant ($X^2 = 801.859$, $p < .001$), indicating the sample was large enough to conduct an EFA (Field, 2009; Henson & Roberts, 2006). Eigenvalues were used to decide the number of factors. Eigenvalues indicated that there was only one factor with eigenvalue above 1.00. It was determined that the FERET has one dimension with 6 items. Furthermore, it was found that one-factor solution of the FERET explains 52% of total variance. Merenda (1997) stressed that the total number of factors should explain at least 50% of total variance, so it was decided that one-factor solution of the FERET is sufficient to explain total variance. Based on the EFA results, it was concluded that the FERET is unidimensional as outlined in Table 4.

Item analysis through IRT

Although traditional item analysis reveals a great deal about test items, the items are vulnerable to influence related to the test takers. IRT produces rich item-level information and has remarkable advantages over classic item analysis. In other words, IRT is useful to identify nonlinear relationships between individual characteristics and their responses to a test (Edelen & Reeve, 2007; Fraley, Waller & Brennan, 2000; Murphy & Davidshofer, 2005). IRT is advantageous in computing standard error of measurement (SEM) for each latent trait so it has more precision (Hambleton & Swaminathan, 1985).

Item characteristics curve (ICC)

The ICC is a basic feature of the IRT, so all other construct and analysis depend on the ICC. The ICC is a graphic representation that indicates the probability of responding with the correct answer and psychological disposition or skill being measured by the test (Edelen & Reeve, 2007; Hambleton & Swaminathan, 1985; Murphy & Davidshofer, 2005). The ICC also provides a large body of information about the relationship between the trait being measured and test responses (Lord, 1980).

In the present study, the graded response model (GRM) was used because the item responses are polytomous (Samejima, 1969). This model reveals the relationship between option parameters, person parameters and the selection of a specific option with data collected with the GRM. Therefore, the ICC for the test items was analysed through the GRM. Graphics of ICC for the test items are displayed in Fig. 1 (Fig. 1, Table 5).

The value of a is the discrimination that indicates differentiation between better respondents and poor respondents based on the latent variable (Edelen & Reeve, 2007; Hambleton & Swaminathan, 1985). Findings indicated that all the items in the model have over 1.00 a value. Based on this finding, all the items can discriminate well between better respondents and poor respondents. B values indicate difficulty level and location of item parameters where respondents have .50 chance of responding correctly to the item. Findings in Table 3 reveal that b values of the items in the model vary between −1.79 and 0.41 (Fig. 2).

Normality tests according to age and gender were conducted to reveal more precise information about homogeneity of the FERET. Normality tests were analysed through skewness and kurtosis coefficient (Table 6).

As a result of the normality test, it was observed that skewness and kurtosis value of all age groups vary between .22 and .48; hence findings indicated that scores of the FERET across the age groups have normal distribution.

Findings from Table 7 indicated that skewness and kurtosis coefficient for male and female participant children vary between .16 and. 34. Therefore, it was decided that the scores of both male and female participant children have normal distributions.

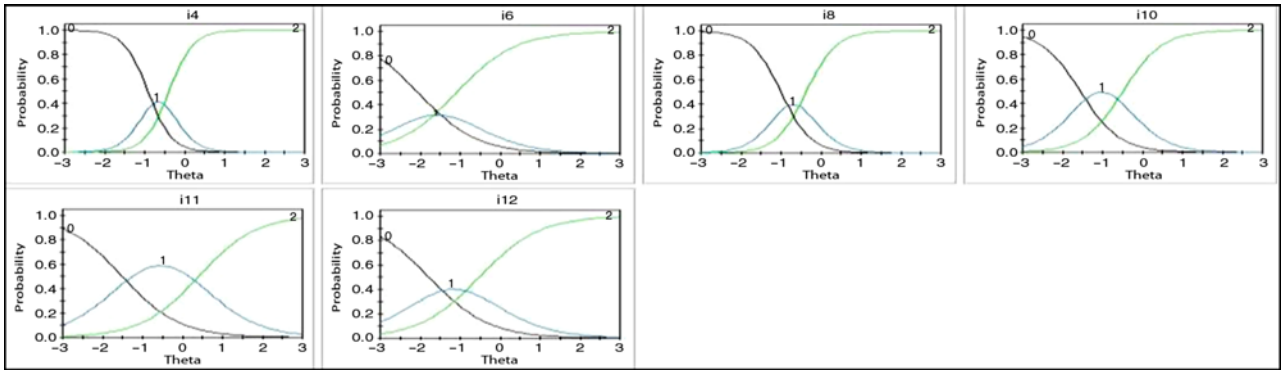

Fig. 1. ICT curves of all the items in the FERET.

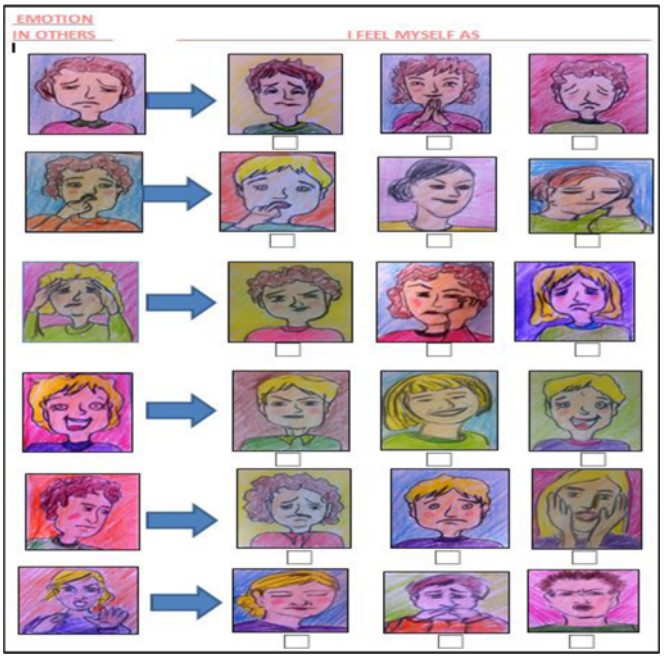

Fig. 2. Final form of FERET.

Discussion

The present study reports on the development of FERET, a tool that measures facial emotional recognition and empathy skills among primary school children. FERET consists of 6 items and one-factor solution. The internal consistency coefficient of FERET was found to be .82. The discrimination values (*a*) for all items were found to be over 1.00, so they differentiate well between poor responders and better responders.

Empathy refers to an emotional response, which is manifested in accordance with recognition of emotions in others (Mehrabian & Epstein, 1972); hence empathy integrates emotion recognition and emotional response to others. Items of the FERET require primary school children to recognise emotional states depicted on faces and choose one of the emotional responses among three response choices, which are theoretically scaled as highly positively correlated, moderately correlated and unrelated–negatively correlated with the recognised emotions. As a result, the FERET involves facial emotion recognition and emotional response, which are components of empathetic skill. Moreover, theoretical scaling based on The Two Factorial Consensual Mood Structure developed by Watson and Tellegen (1985) makes the FERET more objective in deciding which response is truer. Theoretically, it can be considered that the child marking the highly positively correlated response is assessed as a good performer because identifying the highly positively correlated response among the three response choices requires accurate emotion recognition and selection of proper emotional response. On the contrary, choosing an unrelated–negatively correlated emotional response can be evidence of either inaccurate facial emotion recognition or indifference to others' emotional state. Therefore, it can be argued that a higher score on the FERET indicates better possession of facial emotion recognition and empathy skills, whereas lower score on the FERET can be evaluated as a sign of low possession of the two skills.

Facial representations of emotions are the same for all cultures and facial muscle actions are universal. For instance, humans express anger through the same facial musculature regardless of their race, gender, culture and other physical–facial characteristics such as skin and hair colour (Ekman & Friesen, 1971, 1988; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). Based on this notion, it was assumed that perception of emotions in the FERET is not influenced from gender, hair and skin colour at the stage of scale construction. At the same time, it can be noted that emotions represented in the FERET are likely to be perceived by primary school children from different nations in the same way as Turkish primary school children recognise them.

There are a lot of instruments assessing either facial emotion recognition or empathy skills. These instruments can be classified into two groups: Self-report and performance based. Some self-report instruments designed to identify empathy include the Index of Empathy for Children and Adolescents (IECA) designed by Bryant (1982), the Social Skills Improvement System (SSIS) developed by Gresham and Elliott (1990), the Differential Emotions Scale (DES-IV) developed by Izard et al. (1974) to measure emotion recognition in children, and Jolliffe and Farrington's Basic Empathy Scale (BES), which assesses empathetic skills in adolescents. Self-report measurements depend on test participants having objective, consistent, sincere and clear insight about the traits being evaluated. However, self-report measurements are at risk of participant exaggeration or displaying a social desirability bias. Respondents may overestimate their skills, knowledge and traits or, on the other hand, because of social desirability they

Table 6. Normality test according to age

Ages	*n*	$\bar{X}$	*SD*	Skewness	Kurtosis
7 Years	111	13.27	3.59	.22	.45
8 Years	102	13.74	3.21	.23	.47
9 Years	97	16.25	2.76	.24	.48
10 Years	112	15.25	2.43	.22	.45

Table 7. Normality test according to gender

Gender	*n*	$\bar{X}$	*SD*	Skewness	Kurtosis
Female	197	14.00	3.37	.22	.45
Male	225	15.40	2.92	.23	.47

may conceal genuine responses in order to avoid blame from others (Northrup, 1997). Moreover, self-report measurements can be less useful with primary-school children because they do not have adequate metacognitive skills, abstract reasoning and objective thinking which are necessary to evaluate themselves consistently and objectively. Metacognitive skills, abstract reasoning and objective thinking prevent social desirability bias. Consequently, self-report measurements can be considered a less dependable way to collect reliable and valid data from primary school children. Therefore, the FERET relies on a performance-based measure that is more likely to produce precise and reliable results.

There are a number of existing performance-based instruments which measure either facial emotion recognition or empathetic skills for different age ranges. The Diagnostic Assessment of Non-Verbal Accuracy (DANVA), developed by Nowicki and Duke (1989), measures emotion recognition of children through performance-based testing traditions. The Southampton Test of Empathy for Preschoolers (STEP), developed by Howe et al. (2008), is a performance-based instrument measuring empathetic skills in preschoolers. The Assessment of Children's Emotional Skills (ACES), designed by Schultz et al. (2004), is a performance-based instrument measuring emotion recognition skills of children between 4 and 8 years of age. And the JACFEE by Matsumoto and Ekman (1988) was developed to measure facial emotion recognition of adults. Like the FERET, these instruments use performance-based measures, but none of them measure both facial emotion and empathetic skills. The FERET is a multifaceted instrument that can measure both of the skills.

Results of the research indicate that the FERET is a reliable and valid instrument due to its internal consistency coefficient and discrimination power to measure facial emotion recognition and empathetic skills among primary school children whose ages range between 7 years and 10 years.

Conclusion

In the school setting where a strong and positive relationship between academic achievement and empathetic skills is desirable, the FERET has advantages for school staff in terms of monitoring development of primary school children and identifying any problems. As researchers have shown, the development of empathy positively correlates with academic achievement (Caprara et al., 2000; Carlo et al., 2003; Costa & Faria, 2015; Eisenberg & Miller, 1987; Feshbach & Feshbach, 1987; Wentzel, 1993). Moreover, measurement is a key part of instructional processes in order to evaluate whether curricular activities have the expected impacts on primary school children. In addition, not only can primary school teachers use the FERET, so too can school counsellors to detect primary school children's problems with social–emotional skills and monitor their interventions.

The results of the study indicate that FERET generates valid and reliable results in assessing facial emotion recognition and empathy skills among primary school children. In particular, the IRT results demonstrated that the construction of the test effectively discriminates children with better facial emotional recognition and empathy skills from those less able to do so. While a limitation of this study is that a test of the FERET's test–retest reliability was not conducted, this is expected to be done in the course of future research. Although the FERET was only trialled in Turkish primary school classrooms, it is recommended that this performance-based test could be trialled in the other cultural contexts.

Author ORCIDs. Kerem Coskun 0000-0002-3343-2112

References

Ainley, J. (2006). Developing interdependence: An analysis of individual and school influences on a social outcome of schooling. *Educational Psychology*, *26*(2), 209–227.

Bar-On, R., & Parker, J. D. (2008). *Bar-On Emotional Quotient Inventory (EQ-i-YV): Youth version*. Toronto, Canada: Multi-Health Systems.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*(1), 106–148.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413–425.

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, *60*(2), 69–83.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science*, *11*(4), 302–306.

Carlo, G., Hausmann, A., Christiansen, S., & Randall, B. A. (2003). Sociocognitive and behavioral correlates of a measure of prosocial tendencies for adolescents. *The Journal of Early Adolescence*, *23*(1), 107–134.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–108.

Costa, A., & Faria, L. (2015). The impact of emotional intelligence on academic achievement: A longitudinal study in Portuguese secondary school. *Learning and Individual Differences*, *37*, 38–47.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Crowley, K. (2014). *Child development*. London, England: Sage Publications.

Danielson, C. K., & Phelps, C. R. (2003). The assessment of children's social skills through self-report: A potential screening instrument for classroom use. *Measurement and Evaluation in Counselling and Development*, *35*(4), 218–229.

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

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, *32*(4), 257–267.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why?. *Trends in Cognitive Sciences*, *10*(10), 435–441.

De Waal, F. B. M. (2008). Putting back the altruism into altruism: The evolution empathy. *Annual Review of Psychology*, *59*, 279–300.

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology*, *13*, 127–133.

Dymond, R. F. (1950). Personality and empathy. *Journal of Consulting Psychology*, *14*(5), 343–350.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modelling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*(1), 5–18.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91–119.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Ekman, P., & Friesen, W. V. (1988). Who knows what about contempt: A reply to Izard and Haynes. *Motivation and Emotion*, *12*(1), 17–22.

Faupel, A. (2003). *Emotional literacy: Assessment and intervention—Ages 7 to 11*. London, England: Nfer Nelson.

Feshbach, N. D., & Feshbach, S. (1987). Affective processes and academic achievement. *Child Development*, *58*(5), 1335–1347.

Feshbach, N. D., & Roe, K. (1968). Empathy in six-and seven-year-olds. *Child Development*, *39*(1), 133–145.

Field, A. (2009). *Discovering statistics using SPSS*. London, England: Sage Publications.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York, NY: McGraw Hill.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*(2), 350–365.

Gini, G. (2006). Brief report: Adaptation of the Italian Version of the Tromsø Social Intelligence Scale to the adolescent population. *Journal of Adolescence*, *29*(2), 307–312.

Goldenberg, I., Matheson, K., & Mantler, J. (2006). The assessment of emotional intelligence: A comparison of performance-based and self-report methodologies. *Journal of Personality Assessment*, *86*(1), 33–45.

Gresham, D. C., & Elliott, S. N. (1990). *Social skills rating system: Manual*. Circle Pines, MN: American Guidance Service.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. New York, NY: Springer.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393–416.

Hightower, A. D., Cowen, E. L., Spinell, A. P., Lotyczewski, B. S., Guare, J. C., Rohrbeck, C. A., & Brown, L. P. (1987). The Child Rating Scale: The development of a socioemotional self-rating scale for elementary school children. *School Psychology Review*, *16*(2), 239–255.

Hoffman, M. L. (2007). *Empathy and moral development: Implications for caring and justice*. Cambridge, England: Cambridge University Press.

Howe, A., Cate, I. M., Brown, A., & Hadwin, J. A. (2008). Empathy in preschool children: The development of the Southampton Test of Empathy for Preschoolers (STEP). *Psychological Assessment*, *20*(3), 305–390.

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Educational and Psychological Measurement*, *71*(4), 617–637.

Inderbitzen, H. M., & Foster, S. L. (1992). The teenage inventory of social skills: Development, reliability, and validity. *Psychological Assessment*, *4*(4), 451–459.

Izard, C. E., Dougherty, B. M., Bloxom, B. M., & Kotsch, W. E. (1974). *The Differential Emotion Scale: A method of measuring the subjective experience of discrete emotions*. Unpublished manuscript. Nashville, TN: Vanderbilt University.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611.

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*, *8*(4), 369–377.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York, NY: Routledge.

Matson, J. L., Rotatori, A. F., & Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: The Matson Evaluation of Social Skills with Youngsters (MESSY). *Behaviour Research and Therapy*, *21*(4), 335–340.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE)*. [Slides]. San Francisco, CA: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.

Matsumoto, D., Keltner, D., Shiota, M. N., O'sullivan, M., & Frank, M. (2008). Facial expressions of emotions. In M. Lewis, M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 211–234). New York, NY: Guilford Press.

Matthews, B. (2006). *Engaging education: Developing emotional literacy, equity and co-education*. Maidenhead, England: Open University Press.

Mayer, J. D. (2001). A field guide to emotional intelligence. In J. Ciarrochi, J. P. Forgas & J. D. Mayer (Eds.), *Emotional intelligence in everyday life* (pp. 3–24). New York, NY: Psychology Press.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Target articles: Emotional intelligence: Theory, findings, and implications. *Psychological inquiry*, *15*(3), 197–215.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy1. *Journal of Personality*, *40*(4), 525–543.

Merenda, P. F. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counselling and Development*, *30*(3), 156–154.

Merrell, K. W. (1996). Social-emotional assessment in early childhood the preschool and kindergarten behavior scales. *Journal of Early Intervention*, *20*(2), 132–145.

Michelsen, L., & Wood, R. (1982). Development and psychometric properties of the Children's Assertive Scale. *Journal of Behavior Assessment*, *4*, 3–13.

Milgram, N. A. (1960). Cognitive and empathic factors in role-taking by schizophrenic and brain damaged patients. *The Journal of Abnormal and Social Psychology*, *60*(2), 219–224.

Murphy, C. O., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications*. Upper Saddle River, NJ: Pearson.

Northrup, D. A. (1997). *The problem of the self-report in survey research: Working paper*. Ontario, Canada: York University, Institute for Social Research.

Nowicki, S. Jr., & Duke, M. P. (1989). A measure of nonverbal social processing ability in children between the ages of 6 and 10. In Paper presented as part of a symposium at the American Psychological Society, Alexandria, VA.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw Hill.

Petrides, K. V., Sangareau, Y., Furnham, A., & Frederickson, N. (2006). Trait emotional intelligence and children's peer relations at school. *Social Development*, *15*(3), 537–547.

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, *38*(5), 784–791.

Rieffe, C., Terwogt, M. M., Petrides, K. V., Cowan, R., Miers, A. C., & Tolland, A. (2007). Psychometric properties of the Emotion Awareness Questionnaire for children. *Personality and Individual Differences*, *43*(1), 95–105.

Rockhill, C. M., & Greener, S. H. (1999). Development of the Trait Meta Mood Scale for elementary schoolchildren. Paper presented at the Biennial meeting of the Society for Research in Child Development, Albuquerque, NM.

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, *48*(5), 845–858.

Rummel, R. J. (1967). Understanding factor analysis. *Journal of Conflict Resolution*, *11*(4), 444–480.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4), 100–107.

Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology*, *16*(2), 371–388.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology*, *33*(6), 906–916.

Strayer, N. (1987). Affective and cognitive perspective on empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 218–244). New York, NY: Cambridge University Press.

Stueber, K. (2010). *Rediscovering empathy: Agency, folk psychology, and the human sciences*. Cambridge, MA: MIT Press.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357–364.

Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, *26*(2), 173–186.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, *9*(4), 478.

■