Design and validation of a Situational Judgment Test of Socioemotional Competences Development in Young People (SCD-Y)

Diseño y validación del Test Situacional Desarrollo de Competencias Socioemocionales de Jóvenes (DCSE-J)

Rodríguez-Pérez, Sara [1] 🆔**, Sala-Roca, Josefina** [2] 🆔**, Doval, Eduardo** [2] 🆔**, & Urrea-Monclús, Aida** [3] 🆔

[1] University of Oviedo (Spain), [2] Autonomous University of Barcelona (Spain), [3] University of Lleida (Spain)

Abstract

This article presents the design and validation of the situational judgment test of socioemotional competence development in young people (SCD-Y), a copyleft psychoeducational instrument for use with young people aged 12 to 18 years. The test presents five scenarios from everyday life which represent different problematic situations. Interviews were held with 117 young people in order to elaborate the different response options. Seven experts assigned response options to dimensions which had been initially proposed at an earlier stage. In addition, 11 experts assigned discrete scores to each response option. Exploratory factor analysis was carried out on data from 1,227 students attending 12 schools. Test-retest reliability was examined using data from 123 students. Convergent validity was tested using a sample of 215 students who also completed the TMMS-24 and STEM_Y tests. The SCD-Y is a promising psychoeducational evaluation instrument. Notably, it was generally accepted and positive evaluated by teachers, with validity and reliability outcomes also being acceptable up to the present date.

Keywords: Situational test; Socioemotional competences; Youth; Social skills; Emotional intelligence.

Resumen

En este artículo se presenta el diseño y validación del test situacional Desarrollo de Competencias Socioemocionales de Jóvenes (DCSE-J), un instrumento psicoeducativo gratuito (copy-left) para jóvenes de 12 a 18 años. El test se compone de 5 historias de la vida cotidiana con diferentes situaciones-problema. Para la construcción de las distintas opciones de respuesta se entrevistó a 117 jóvenes. La asignación de las diferentes respuestas a las dimensiones planteadas inicialmente la realizaron 7 personas expertas. Además, 11 personas expertas asignaron puntuaciones discontinuas a cada opción de respuesta. Resultados: El análisis factorial exploratorio se llevó a cabo a partir de los datos de 1227 estudiantes en 12 centros educativos. La prueba test-retest se comprobó con 123 estudiantes. La validez convergente fue comprobada con 215 estudiantes que también respondieron al TMMS-24 y al STEM_Y. El DCSE-J es un instrumento de evaluación psicoeducativa prometedor. Destaca la aceptación y valoración positiva por parte del profesorado y los resultados de las pruebas de validez y fiabilidad llevadas a cabo hasta el momento.

Palabras clave: Test situacional; Competencias socioemocionales; Juventud; Habilidades sociales, Inteligencia emocional

Corresponding author / Autor de contacto: Sara Rodríguez Pérez. University of Oviedo. Faculty of Education.
C/Aniceto Sela s/n, 33005, Oviedo, Asturias (Spain). correo-e: rodriguezpsara@uniovi.es

Over a century ago, Darwin (1872) noted that children are born with basic emotions that ensure their survival. This range of emotions was widened and diversified with the identification of self-consciousness and social interaction (Lewis et al., 1989). However, not all children are equally predisposed to experiencing the same emotions, nor do they experience them with equal intensity. These differences are inherent to temperament, which is transferred genetically and, subsequently, influences and is influenced by individual experiences, weaving the pathway to adult personality (Rothbart et al., 2000).

In the same way as seen in other species, the first human emotions were identified through phylogenesis and their emergence represented a breakthrough in adaptative pathways by providing complex responses to critical situations in order to support survival of the individual and the species. Such emotions included responses such as flight from a predator or the protection of a baby (Darwin, 1872). However, human interactions are central in today's societies and primary emotional responses often do not facilitate social adjustment. The complexity of social interactions requires skills that mediate the emotions these relationships generate in us and the responses we have to these situations. In recent decades, research has agreed that both expressing emotions (Hernández, et al., 2016), and understanding one's emotions and those around us are key to relating with others (Kar & Kar, 2017; Laghi et al., 2018), while many psychopathological disorders have been found to be rooted in the difficulty of regulating emotions (Sheppes & Gross, 2015).

Emotional skills develop from the early stages of childhood, stimulated by social interaction (Grusec, 2011). This development involves a process in which emotional and social skills intertwine, making it impossible to disaggregate them (Monnier, 2015; Boyatzis et al., 2015). It also takes place at different tempos; thus, for example, emotional understanding develops early, while regulation is usually mastered at more advanced stages with the maturation of frontal lobes during adolescence facilitating emotional regulation

skills (Blakemore & Mills, 2014; Steinberg, 2005). This being said, the level of competence in these skills may increase throughout the lifespan. In this regard, Bar-On observed that the highest scores in emotional and social competence tend to be achieved between the late forties and early fifties (Bar-On & Parker, 2000).

Various models of emotional intelligence have been developed. Two accepted models refer to ability models and trait models. The former considers emotional intelligence to comprise a set of cognitive abilities that facilitate information processing and emotional management. These models often employ optimal performance tests which have been traditionally used to measure cognitive intelligence. The latter considers emotional intelligence to be a set of skills, competences and non-cognitive skills which allow us to deal with environmental demands successfully. They typically employ self-report tests, which are also commonly used to assess personality. The data obtained to date would indicate that the two types of model are measuring different constructs given the low levels of intercorrelation between them (Matthews et al., 2017).

Adopting the emotional competences approach, Saarni (2008) underlined the interconnection between social context and emotional development. In this regard, she defined emotional competence as one's effectiveness when it comes to emotionally eliciting social transactions (Saarni, 1997). Thus, as children learn emotional competences in different contexts, they demonstrate emotional competence through their behavior (Saarni, 2000). We agree with Saarni's approach, which deems the concept of socio-emotional competence to be more useful in the educational field. School education aims to educate children in an integral way by promoting their positive social insertion and well-being, an outcome that will not be possible without education on socio-emotional skills. In this context, competence offers a pedagogical framework that is richer than that of skills, because, as Everwijn et al. (1993) have

indicated, the ultimate goal of education is to teach students to apply what they learn.

In line with this, Lievens and Chan (2017) indicated that three types of intelligence are involved in carrying out or executing a task: emotional, social and practical - all of which are multidimensional. These intelligences are believed to have a partial mediating effect on individual characteristics and factors (knowledge, skills, motivations, experience, values, cognitive styles, personality traits, etc.). Within this conceptual framework, situational tests are more effective than other types of test as they are contextualized and have better predictive validity.

In consideration of that discussed above, together with the manifest need to attend to emotional education, the evaluation of socio-emotional skills and/or emotional intelligence (EI) has been one of the main interests in the socio-educational and psychological professional field since the 1990s. This becomes evident if we analyze the growth of scientific productivity in this area (Ciarrochi et al., 2006). However, few instruments can be found that have adequate psychometric properties to perform the evaluation of these competences in childhood and adolescence, whilst those that do exist are cost-prohibitive (Rodrigo, 2017; Sánchez-Teruel & Robles-Bello, 2018).

Self-report instruments have traditionally played a predominant role in measuring socio-emotional skills and emotional intelligence (Bar-On, 1977; Salovey et al., 1995; Mayer et al., 2002a y b). However, these types of measures have been heavily criticized in recent decades. There is, therefore, a need to develop instruments that consider the effects of response formats and scoring methods on the results obtained from tests (MacCann & Roberts, 2008; Allen, et al., 2014). Situational judgement tests (SJT) are thought to meet these criteria given that they presenting hypothetical situations alongside alternative response options from which the individual being evaluated must choose. These tests, therefore, obtain information on the approach considered by individuals to be most likely to achieve intended objectives, providing them with opportunities to offer solutions to the problem posed (MacCann et al., 2010; Allen et al., 2015; Rodrigo, 2017).

Regarding SJT instruments, several authors have highlighted their potential for measuring various competences, although they are still infrequently used instruments and more research is required to support the metric properties of their scores (MacCann & Roberts, 2008; Sharma et al., 2013; O'Connor et al., 2019).

It is also worth noting that, in the educational field, emotional education programs must be accompanied by an evaluation of the target group's competences in order to adapt these programs to the developmental stage and specific needs of each group. In this regard, instruments must be adapted to different ages in order to provide valid, reliable and change-sensitive scores (Bisquerra & Pérez-Escoda, 2015). However, there are few free-to-access instruments available to education professionals for conducting psychoeducational evaluations of young people's socio-emotional skills.

Given the above, it was considered pertinent to embark upon the construction and validation of a free-access instrument based on situational judgement aimed at young people aged 12 to 18. In the present article, the design of a test to assess socioemotional competences in youth is described, alongside the outcomes of three studies which evaluated different validity indicators. The first of these studies aimed to analyze the factor structure of the test, whilst the second and third studies evaluated temporal stability and convergent validity, respectively.

Elaboration of the situational judgment test of socioemotional competence development in young people (SCD-Y).

Below, the process through which the situational judgment test of socioemotional competence development in young people (SCD-Y) was elaborated is described. The test aims to examine the development of young

people's socio-emotional skills and self-esteem based on the responses they give to different daily life scenarios. It is a copy-left psychoeducational instrument designed for use by the educational community which aims to provide data to support evaluation and intervention in different educational (school and non-school) environments.

The procedure described for the creation of situational judgment tests by Lievens, Peeters and Schollaert (2008) was used as a reference to develop the SCD-Y situational test. The aforementioned authors proposed a procedure to develop situational judgement tests for personnel selection. This involved, first, selecting workplace situations that are representative of the competences to be evaluated and the scenarios in which they are employed. Next, expert and non-expert responses to these scenarios were collected from employees.

However, although the procedure suggested by these authors is suitable for occupational environments, it is difficult to apply to a competence development test such as the SCD-Y. The procedures was, therefore, slightly modified as described below in order to adapt it to the restrictions inherent to this type of test.

Phase 1. Design of test scenarios

Dimensions

The situational judgment test of socioemotional competences in young people (SCD-Y) was elaborated in consideration of five competences necessary for regulating behavior and adjusting to the social environment. Namely, these competences were understanding one's own emotions, understanding the emotions of peers (empathy), regulating one's emotions, regulating the emotions of peers and assertiveness. A scale for measuring self-esteem was also included, which, although not a competence, is an emotional dimension that affects social interactions and interpretations of these interactions.

Dimensions were selected considering the areas commonly conceptualized in psychological development manuals, which

recognize the close interrelation of emotional and socio-emotional competences. Four discussion groups with five expert academics in emotional development and education were held to agree on the competences and competence elements that should be considered in the test dimensions.

Contexts

Situational tests provide greater ecological validity (Lievens & Chan, 2017) because responses are given in relation to specific contexts and situations. This overcomes the limitation of self-report tests, which require participants to give general assessments of an aptitude or characteristic without specifying the context in which it occurs. Situational tests assume that the level of execution may vary depending on the problem or context. Hence, the contexts in which the behaviors or competences to be evaluated are observed must be defined during elaboration of the test. In this respect, the execution of socio-emotional competences will also vary depending on the context and problem faced by the adolescent.

The different spaces of socialization experienced by 12 to 18-year-old adolescents were considered when designing the test. These were divided according to the three most common daily life contexts experienced by individuals at these ages: family, friends and school. Five common situations were designed that are routinely faced by adolescents: conflicts with friends, working in a group, changes to new environments, going to parties and family arguments. Despite not all young people necessarily having had direct experience of them, these situations evoke emotions known to be common at this phase of evolutionary development. Such emotions include frustration, anger, disappointment, shame, etc. They also pose challenges to young people on a daily basis; for example, defending oneself, understanding one's own emotions and those of others, regulating emotional processes, etc.

Creating problem situations

Defining the problem situations in which the behaviors or competences to be measured may be observed constitutes a fundamental element of situational tests. Many such tests pose a

different situation for each item. However, situating oneself in different contexts involves significant cognitive effort and could generate fatigue among participants, especially adolescents. The scenarios chosen in the design of the SCD-Y were, therefore, naturally occurring situations in order to make it easier

for participants to immerse themselves in the proposed context without great cognitive effort.

In total, five daily life scenarios were proposed. Each scenario included six different situations, with one to evaluate each of the six dimensions under study (the five competences plus self-esteem) (Table 1).

Table 1. Test scenarios

Context	Scenario	Number of situations
Friends	Where to go for a day out	6
School	Group work	6
Family	I've moved to a new city	6
Friends	The party	6
Family-school	They won't buy me what I've asked for	6

The gender perspective and social exclusion

In order to help adolescents identify with the characters in the test scenarios, scenarios were attributed with similar characteristics. Thus, the gender of characters was considered, with the same number of males and females appearing in the different scenarios and roles also being alternated.

Furthermore, in the final versions of the test, scenarios were reviewed in which references were made to specific family members. Many adolescents in care do not live with their parents because they are in residential homes, or live with other relatives in foster care, often separate from their siblings. Although such individuals could respond to the test as they could relate to some of the situations, references to family were deemed to tap into a sensitive matter. Thus, references to parents and siblings were replaced with more general expressions (e.g. "you are told that", instead of "your parents tell you that"), thus facilitating the representation of a greater diversity of families and/or ways of life.

Phase 2. Elaboration of response options

In order to elaborate the most representative response options for the 30 situations constituting the test, semi-structured interviews were conducted with a group of 117 young

people (36 boys and 81 girls) aged between 12 and 18 (12-13-year-olds = 31; 14-15-year-olds = 34; 17-18-year-olds = 52). In the interviews, respondents were presented with the five scenarios and their respective six situations. Responses were transcribed and analyzed by means of content analysis and a deductive-inductive mixed categorization process was then carried out.

Based on the analysis of interviews, five response options were selected to represent different ways of reacting to or acting in response to presented situations. To this end, it was considered that: a) responses should reflect different levels of competence, and b) responses should have been given by several young people in order to ensure variability.

Phase 3. Elaboration of questions to measure response consistency

Initially, two questions were included in order to detect individuals who answered randomly or dishonestly, or who may have had problems understanding the test. Specifically, participants were asked whether they had answered at least some of the questions honestly or randomly. Subsequently, when examining reliability (using the test-retest method), these two questions were replaced with three items that replicated one of the test

items and were located in different scenarios to those of the original item. The response options were the same as in the original item, although the order in which they were presented differed.

Phase 4. Expert scale validation

Once the test had been designed, its structure was validated based on criteria defined by seven experts. The expert group comprised five females and two males who were professionals (3) or scholars (4) in the fields of education (5) and psychology (2).

Each expert was asked to identify which of the six dimensions of the test was required to adequately address each of the 30 proposed situations. Thus, the number 1 was assigned to the selected dimension and 0 to non-selected dimensions. On the basis of this, discrepancies between judges were examined in accordance with the median scores for each situation. Average discrepancy scores ranged from 0.01 to 0.12.

In cases where discrepancies were found between two or more of experts and kappa coefficients did not exceed 0.80 (Cohen, 1960), corresponding situations were analyzed individually. For this, all experts gathered together to analyze the situations and adjust the wording of the questions and/or response options.

Phase 5. Assigning scores to response options

In order to assign scores to the different response options, a procedure proposed by McCann (2008) was followed in which experts are requested to assign scores.

Thus, 11 experts (six men and five women) from the field of socio-emotional competences (professionals in psychology and psychopedagogy) were asked to assign a score of between 0 and 5 to each of the options. Responses corresponding to a total lack of competence were assigned a score of 0, whereas, responses corresponding to maximum competence were assigned a score of 5.

Agreement between judges was analyzed by calculating the distance of assigned scores with respect to the median score assigned by the overall group of experts. Mean discrepancy was calculated and two experts were found to have assigned scores that exceeded the median by an average distance of one. The scores of these experts were excluded from analysis, leaving a final sample of 9 experts (four male and five female).

In addition, nine degree students (one male and eight females studying primary education, pedagogy and social education) were invited to participate in the process. The aim of this was to examine whether generational differences affected scoring. These students followed the same process described above and none of the scores provided exceeded the median by more than one. High inter-rater agreement ($\kappa = .92$) was confirmed between the two groups of evaluators using the kappa index.

In order to clarify and compare quantitative data, a focus group was conducted with the above students in order to analyze differences between the scores assigned by experts to the consistency items. This analysis focused on one of the questions. In the focus group, non-equivalence was deemed to be present in one of the situations, which was subsequently adjusted to ensure equivalence and the effectiveness of consistency control mechanisms.

As a result of the above procedure, the scores assigned to each item (Table 2) corresponded to the median scores given by the 18 professional experts and students, indicating good inter-rater agreement ($\kappa =.86$). This test can be consulted online (Sala-Roca et al., 2016).

Phase 6. Pilot test

A pilot test was carried out with 36 young people aged between 12 and 18 years who were relatives and acquaintances of research group members. The aim of this was to examine understanding of the test, the time required for completion, fatigue and response variability.

Table 2. Scores assigned by experts to each of the scenario-problem situations in order to measure socio-emotional skill level

	Scenario 1: Where to go for a day out?	Scenario 2: Group work	Scenario 3: I've moved to a new city	Scenario 4: The party	Scenario 5: They won't buy me what I asked for
Self-esteem	Ex1 (0 3 0 5 1)	T1 (0 5 3 2 4)	Ci1 (1 1 4 5 5)	F2 (5 3 1 0 4)	Co1 (0 2 0 5 4)
Assertiveness	Ex2 (0 4 5 0 2)	T6 (2 5 3 0 0)	Ci5 (5 4 1 1 0)	F1 (5 0 0 3 1)	Co3 (4 1.5 0 0.5 4)
Understanding the emotions of others	Ex3 (3 4 3 1 0)	T2 (3 3 5 0 0)	Ci3 (0 4 4 1 0)	F5 (4 0 3 3 0)	Co5 (4 3 3.5 2 0)
Understanding one's own emotions	Ex4 (3 2 4 3 0)	T4 (3 3 4 4 0)	Ci4 (3 4.5 3.5 3 0)	F4 (4 5 1 2 0)	Co4 (3 4.5 2.5 2 0)
Self-regulation	Ex5 (0 1.5 0 5 4)	T5 (0 0 4.5 4 3)	Ci6 (5 2 3 0 0)	F3 (3.5 4 0 3.5 0)	Co2 (0 2 4 5 4)
Emotional regulation of others	Ex6 (4 0 5 4 1)	T3 (3 0 3 4 0)	Ci2 (4 3 5 0.5 4)	F6 (4 5 0 0 4)	Co6 (4 5 0 0 3)

Note: Ex1-6 (items 1-6 from the scenario "Where to go for a day out?"); T1-6 (Group work); C1-6 (I've moved to a new city); F1-6 (The party); Co 1-6 (They won't buy me what I asked for).

No problems with understanding test questions were detected. Response time ranged from 25 to 45 minutes. Participants rated the test to be highly enjoyable. All response options were selected for 51 of the 60 test items. One of the given response options was only not selected in response to nine of the items. As a result, response variability was deemed to be acceptable.

Phase 7. Administration

Subsequently, the test described above was administered to 1,227 students attending 12 schools in order to obtain the necessary data to perform exploratory factor analysis (see study 1). Following administration, a focus group was conducted with participants' teachers to analyze the way in which the procedure had been carried out, administration times, difficulties with understanding and the degree of interest shown by students.

Outcomes indicated that administration times ranged from 30-45 minutes, few problems understanding the test were experienced (with those reported being mostly limited to the understanding of a single word), and students identified well with the different scenarios and situations, showed great interest in the results and rated the test as an enjoyable and useful activity.

Phase 8. Convergent validity analysis

In order to analyze convergent validity, a group of 215 students were administered the SDC-Y test, together with two other tests designed to measure aspects of socioemotional skills, namely, the TMMS-24 and the STEM_Y. The former is an adaptation of the trait meta-mood scale devised by Salovey et al. (1995) and adapted by Fernández-Berrocal et al., (2004). This test was designed to assess perceived intrapersonal emotional intelligence (emotional attention, emotional clarity and emotional repair) and asks respondents to report their level of agreement with statements about emotional experiences without specifying the context. The STEM_Y test (MacCann et al., 2010) is a shortened version of 14 questions designed for use with the adolescent STEM population. It is a situational test that measures emotional regulation in different contexts. It was adapted into Spanish by Rodríguez-Ruiz (2017), reducing the number of items to 11. Outcomes provided evidence of convergent validity (see study 3).

Study 1: Factor structure

Sample

12 secondary schools who expressed interest in the study were recruited via von-probabilistic sampling (convenience

sampling). A total of 1227 students completed the test. Tests indicating random or dishonest responses were eliminated. This left a total of 932 participants (50.6% boys and 49.4% girls) aged between 11 and 18 years ($M = 14.1$, $SD = 1.82$).

Procedure

School directors were informed about the study aim, procedure and data protection protocols. After schools provided informed assent, the procedure was presented to the teachers who would collaborate in data collection. Schools informed families and students about the study goals and data protection procedures and requested voluntary informed participation. School tutors administered the test electronically. In order to preserve anonymity, students entered a reference number when completing the test which identified them and alerted their tutors to their participation. No personal identifying data were collected. Results were sent to tutors to pass on to their students. Didactic activities for teachers and parents were also provided to work on weaknesses highlighted by test outcomes with the aim of improving socioemotional competences.

Data analysis

Only three items showed high kurtosis (C2, C4 and D1) and one item had pronounced asymmetry (D1). This being said, none of the items presented a normal distribution according to the Kolmogorov-Smirnov test. Unrestricted factor analysis (Ferrando & Anguiano-Carrasco, 2010; Ferrando & Lorenzo-Selva, 2000) was performed using the FACTOR 10.3 program (Lorenzo-Seva & Ferrando, 2007, 2011). The unweighted least square method with oblimin rotation was used. The following goodness-of-fit indices were considered: chi-square significance test, Tucker Lewis Index or non-normed fit index (TLI-NNFI), comparative fit index (CFI), goodness of fit index (GFI) and root mean square residuals (RMSR). Goodness of fit is indicated when TLI-NNFI values are close to .90 (Byrne, 2001), CFI values are close to .95 (Hooper et al., 2008) and SRMS values are below .08 (Byrne, 2001).

Internal consistency and reliability of scores over time were analyzed using the SPSS v22 program for the scales resulting from each factor.

Results

Several factor models were analyzed. Table 3 presents goodness-of-fit indices. Values pertaining to the KMO (.80) and the Bartlett statistic (3261.2, $df = 435$, $p < .001$) were acceptable. Initially, a six-factor model was tested, with factors pertaining to each dimension proposed by the test (model 1). Goodness-of-fit indices were generally acceptable. However, a coherent interpretation of item distribution was not possible given that some factors were underrepresented (three factors were formed by only two items) and nine items were not included in any factor (three of which corresponded to situations related to self-esteem). The decision was therefore made to discard this factor structure.

For the next model (model 2), a three-factor structure was considered by grouping together items referring to regulation (of oneself and of others) and assertiveness in one factor, items relating to understanding (of oneself and others) in another, and items corresponding to self-esteem in a third. This model presented worse fit than the first, with acceptable GFI and SRMR values but lower CFI and TLI values. As with Model 1, item distribution between the different factors could not be interpreted. In addition, six of the items had loadings lower than .30 on all factors (of which, four corresponded to self-esteem-related items). Thus, the decision was made to also discard this factor structure.

For the third model (model 3), it was decided to test a two-factor model, grouping together items related to regulation and assertiveness, on the one hand, and items referring to understanding, on the other. It was decided not to include items pertaining to self-esteem since these produced the weakest indices in the previous two models. As in the case of model 2, GFI and SRMR values were acceptable but CFI and TLI values were low. As can be seen in table 4, the factor pertaining to emotional understanding was defined by eight items with

loadings greater than .30, whilst the factor pertaining to emotional regulation and assertiveness was defined by 14 items, of which 11 had loadings greater than .30 and three (A6, B3, E3) had loadings close to this value. Three of the test items produced loadings greater than .30 for both factors. The first factor, emotional comprehension, presented a Cronbach's alpha of .65, while this value was .76 for the second factor (regulation-assertiveness). A significant correlation ($r = .51$, $p < .001$) was found between the two factors.

Table 3. Model fit

Model	n	χ^2	df	CFI	TLI	GFI	SRMR
Model 1	932	3261.177 ***	435	.95	.92	.99	.02
Model 2	932	3261.177***	435	.87	.83	.98	.04
Model 3	932	681.902***	300	.82	.78	.97	.04

* p < .05; ** p < .01; *** p < .001

Table 4. Items of the SCD-Y

Item	Factor	Factor loading	Mean	Standard Deviation
B6	R-A	.49	2.23	1.80
E6	R-A	.49	3.33	1.90
B5	R-A	.47	2.16	2.27
C6	R-A	.44	3.12	1.81
C5	R-A	.41	2.95	1.97
D3	R-A	.41	3.12	1.65
A5	R-A	.40	2.63	1.92
C2	R-A	.38	3.75	1.11
D4	R-A	.35	3.68	1.77
A2	R-A	.30	2.76	1.78
E2	R-A	.30	2.04	1.54
E3	R-A	.28	3.34	1.28
A6	R-A	.27	3.21	1.58
B3	R-A	.26	2.62	1.44
D6	R-A / U	.45 / .37	4.06	1.74
D1	R-A / U	.49 / .36	3.78	1.78
C3	R-A / U	.32 / .31	3.35	1.44
B4	U	.44	3.05	1.50
D5	U	.41	3.15	1.39
E5	U	.39	3.02	1.32
B2	U	.37	3.16	1.77
C4	U	.36	3.38	1.18
A4	U	.34	2.99	1.20
E4	U	.32	3.05	1.82
A3	U	.31	2.44	1.48

Note. R-A = Emotional regulation-assertiveness; U = Emotional understanding

Finally, it should be noted that three of the items were considered to belong to two factors since loadings greater than 0.3 were produced in relation to both (items C3, D1 and D6). This is consistent with the interrelated nature of socio-emotional competences. Emotional regulation requires an understanding of emotions. This interrelated nature implies that the two factors have significant correlations ($r = 0.48$, $p < 0.005$).

Study 2: Test-retest reliability

Sample

Non-probabilistic sampling (convenience sampling) was used to recruit university students. A total of 123 students participated in the study voluntarily. These students were studying degrees on early childhood education (49.7%), primary education (10.9%), social education (33.6%) and pedagogy (7.6%) in the Faculty of Education Sciences at the Autonomous University of Barcelona. Following analysis of questionnaire consistency, data pertaining to scores higher than 6 were excluded, leaving a final sample of 119 students, of which 107 (89.9%) were female and 12 (10.1%) were male. The average age was 20.9 years ($SD = 3.46$).

Procedure

The SCD-Y was administered online using the Google Docs tool. The procedure was as follows:

- Lecturers of the selected university degrees explained the aim and importance of the research to students. Students were informed about data confidentiality and the way in which obtained data would be handled. Voluntary participation was requested.
- A time-frame of one week was given to complete the instrument.
- Once three weeks had lapsed, students were sent an email reminder to complete the questionnaire.

Data analysis

Data analysis was carried out using version 22 of the SPSS statistical program. Pearson correlations were calculated.

Results

Positive correlations were found between the scores obtained for each of the two factors in the first and second administration of the questionnaire (Table 5). As expected, no correlation was found between regulation-assertiveness and understanding.

Table 5. Correlations between factor scores obtained in the first and second questionnaire administration

First administration	Second administration	Correlation
Regulation-assertiveness	Regulation-assertiveness	.682**
Understanding	Understanding	.617**
Regulation-assertiveness	Understanding	.043
Understanding	Regulation-assertiveness	-.041

** p<.001

Study 3: Convergent validity

Sample

Secondary schools interested in the study were recruited via non-probabilistic sampling (convenience sampling). 215 students (43% girls and 57% boys) aged between 12 and 18 participated in the study. Of these, 159 were in lower secondary education and the other 56 were in upper secondary education or vocational training.

Procedure

The same procedure outlined for study 1 was used to contact and inform management, teachers, students and families. The SCD-Y, TMMS-24 and STEM-Y were administered to students in a single session. Students' tutors were in charge of administering the tests. In order to protect anonymity, tests were identified using reference numbers which only students and tutors had access to.

Data analysis

Data analysis was carried out using version 22 of the SPSS statistical program. Pearson correlations were calculated between the scales of the different tests, and between scores and ages. In addition, means were compared to analyze whether differences varied according to gender.

Results

SCD-Y outcomes were positively correlated Stem Y test outcomes. The two general SCD-Y factors of emotional regulation and understanding were also positively correlated with this test ($r = .34$ and $r = .35$, respectively; $p < .001$).

No correlations were observed between the two SCD-Y factors and the three TMMS-24 scales, nor were correlations observed between TMMS-24 scale and Stem_Y test outcomes.

Gender differences were observed in some test scores. Boys scored slightly lower than girls on the emotional understanding factor of the SCD-Y (boys: $M = 97.4$, $SD = 15.9$; girls: $M = 103.5$, $SD = 12.9$; $t = -$, $p = .003$). However, higher scores were observed in boys for two of the three TMMS-24 scales, specifically, with regards to emotional clarity (boys: $M = 25.2$, $SD = 6.8$; girls: $M = 22.5$, $SD = 6.6$; $t = 2.2$, $p = .03$) and emotional repair (boys: $M = 25.3$, $SD = 6.6$; girls: $M = 22.8$, $SD = 6.4$; $t = 2.1$, $p = .04$). No gender-related differences were observed in relation to Stem-Y test outcomes.

Age correlated positively, although very weakly, with the regulatory factor ($r = .16$; $p = .03$) of the SCD-Y and the regulatory scale of the other examined test ($r = .14$; $p = .046$). In contrast, negative correlations were observed with the TMMS test for the emotional clarity ($r = -.25$, $p = .01$) and emotional repair ($r = -.19$, $p = .04$) scales. No correlations were observed between Stem-Y outcomes and age.

Discussion and conclusions

Socio-emotional competences influence a multitude of circumstances in daily life, with high emotional intelligence scores having been found to be associated with higher scores in self-esteem, happiness, mental health and life satisfaction, and lower scores in anxiety, depression and the suppression of negative thoughts (Extremera, 2003).

One of the great challenges in measuring these competences stems from their situational nature. In other words, measurement must take place in relation to a context that is familiar to people. In fact, some authors, such as Saarni (2008), consider it inconceivable to separate emotional competence from the social context. Emotional competences are not only learned principally in social contexts but are also necessary for dealing with most social transactions. Thus, once the lack of valid and reliable instruments for their measurement had been detected, it was decided to develop a situational instrument to address this issue. This was intended to overcome the difficulties posed by self-report tests, in which participants are often requested to provide decontextualized responses. Situational tests can also be administered to groups and assess various competences that are difficult to address with ability-based tests. In line with this, the SCD-Y test proposes contexts that students can easily identify with, while the structure of items developed in the short scenarios makes administration enjoyable and facilitates attention and identification. This test is proposed as a psychoeducational instrument. It not only allows for a psychoeducational diagnosis to identify competence levels and facilitate the planning of training interventions but is also an educational instrument in its own right. Teachers can use it to analyze proposed situations with their students, the similarity of these situations with others that they may be experiencing and their responses to them.

Satisfactory evidence was obtained in relation to test content, that is, its operationalization through different scenarios, questions and responses since expert assessments presented high levels of inter-rater agreement (18), regardless of the diversity of profiles included (professionals/non-professionals, different genders and ages).

Furthermore, with regards to its internal structure, exploratory factor analysis was

carried out, obtaining two interpretable factors defined by items with high loadings on the corresponding factors, with the exception of three items, which obtained lower scores (A6, B3, E3). Thus, the solution was interpretable even when TLI and CFI indices were low with regards to the model fit. For this reason, outcomes should be considered with caution and more evidence must be collected in this respect.

Although fit outcomes could be improved, from an interpretative point of view the factor model was acceptable, considering that the six initially proposed theoretical factors are not independent. The difficulty in finding an acceptable factor model could be explained by the characteristics of the situational tests (Sorrel et al., 2016) and also by the scores associated with the response format, which did not correspond to the same scales (Frias-Navarro & Pascual-Soler, 2012).

Despite what might have been expected given the characteristics of this type of test (Sorrel et al., 2016), Cronbach's alpha outcomes pertaining to the two identified scales were acceptable, although not high. Reliability over time could also be considered to be acceptable. In addition, the high ecological validity of the test (Kvavilashvili & Ellis, 2004) should also be considered. This property is closely linked to the representativeness of the test's content, and, therefore, the correspondence between situations proposed by the test and experienced in real life, with this permitting greater generalization of findings. Given that situational tests use items that reflect real everyday environments they better represent reality (Lievens & Chan, 2017). In the SCD-Y test, included scenarios and items reflected everyday situations facing adolescents.

The results obtained in study 3 show moderate or weak correlations that could also support the validity of the construct. The fact that the correlations were moderate or weak may be the result of that fact that the skills measured in the three tests were not exactly the same. Indeed, the TMMS-24 was designed to assess perceived emotional intelligence, whilst the SCD-Y aims to measure socioemotional skills.

On the other hand, the weak correlation could also be a product of the different test formats used. Berrocal & Extremera (2006) has previously highlighted weak correlations between data obtained in self-report tests and performance tests. In fact, the correlations between SCD-Y and STEM_Y outcomes were stronger than those produced between SCD-Y and TMMS-24 data, with no correlation being observed between STEM_Y and TMMS-24 data. Decontextualized assessments of personal emotional experiences are likely to be influenced by social desirability and the tendency to report positively or negatively about oneself. In fact, it is surprising that boys scored higher on emotional clarity in this test, with this potentially indicating a greater effect of social desirability. Sánchez et al. (2008) found that men report higher emotional intelligence than they actually demonstrate on performance tests, whilst the opposite is seen in women who often perform better on performance tests. The SCD-Y was weakly correlated with age, a finding which has also been reported in other studies (Fariselly et al., 2006; Fernández-Berrocal, et al. 2012). Thus, the negative correlation between the TMMS-24 and age is surprising. These outcomes likely have more to do with social desirability and difficulties in making judgments about one's own behavior in the absence of contextualization.

Based on the results of the present study, the SCD-Y test can be considered to be a valid tool for assessing the emotional development of children between the ages of 12 and 18. This psychoeducational tool may be useful when designing interventions tailored to the needs of this group.

Research limitations

Whilst outcomes of the present study are promising, some limitations should also be highlighted. A relatively small sample size was used in studies 2 and 3 which may have affected outcomes. Furthermore, convenience sampling (non-probabilistic sampling) was employed. Finally, factor analysis is limited when applied to situational tests.

Future perspectives

The limitations present in this study should be addressed in future research. In this sense, it would be interesting to expand to broader samples and perform confirmatory factor analysis. Similarly, criterion validity of the tool should be examined. Further examination of convergent validity could also be of interest. Finally, future work should analyze the effect of age and/or gender on outcomes.

Funding and Acknowledgments

References

Allen, V., Rahman, N., Weissman, A., MacCann, C., Lewis, C. y Roberts, R. D. (2015). The Situational Test of Emotional Management – Brief (STEM-B): Development and validation using item responsive theory and latent class analysis. *Personality and Individual Differences, 81*, 195–200. https://doi.org/10.1016/j.paid.2015.01.053

Allen, V. D., Weissman, A., Hellwig, S., MacCann, C. y Roberts, R. D. (2014). Development of the situational test of emotional understanding – Brief (STEU-B) using item response theory. *Personality and Individual Differences, 65*, 3–7. https://doi.org/10.1016/j.paid.2014.01.051

Bar-On, R. y Parker, J. (2000). EQ-I: YV. *Bar-On Emotional Quotient inventory: Youth version. Technical manual*. Multi Health Systems, Inc. https://doi.org/10.1037/t14077-000

Bisquerra, R. y Pérez-Escoda, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *REIRE, Revista d'Innovació i Recerca en Educació, 8*(2), 129-147. https://doi.org/10.1344/reire2015.8.2.82

Blakemore, S. J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual review of psychology, 65*, 187-207. https://doi.org/10.1146/annurev-psych-010213-115202

Boyatzis, R. E., Gaskin, J., & Wei, H. (2015). Emotional and social intelligence and behavior. *In Handbook of intelligence* (pp. 243-262). Springer. https://doi.org/10.1007/978-1-4939-1562-0_17

Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46. https://doi.org/10.1177/001316446002000104

Ciarrochi, J., Forgas, J.P. & Mayer, J.D. (2006). *Emotional Intelligence in everyday life.* Psychology Press/Taylor and Francis.

Darwin, Ch. (1872). *The Expression of Emotion in Man and Animals*. Appleton Century Crofts. https://doi.org/10.1037/10001-000

Extremera, N. (2003). *El modelo de inteligencia emocional de Mayer y Salovey y su validez predictiva en muestras españolas*. Málaga. https://doi.org/10.21840/siic/159626

Everwijn, S. E. M., Bomers, G. B. J., & Knubben, J. A. (1993). Ability-or competence-based education: bridging the gap between knowledge acquisition and ability to apply. *Higher education, 25*(4), 425-438. https://doi.org/10.1007/bf01383845

Fariselli, L., Ghini, M., & Freedman, J. (2008). Age and emotional intelligence. Six Seconds: *The Emotional Intelligence Network*, 1-10.

Rui-Berrocal, P., Cabello, R., Castillo, R., & Extremera, N. (2012). Gender differences in emotional intelligence: The mediating effect of age. *Behavioral Psychology*, 20(1), 77-89.

Fernández-Berrocal, P., & Extremera, N. (2006). Emotional intelligence: A theoretical and empirical review of its first 15 years of history. *Psicothema*, 18, 7-12.

Fernández-Berrocal, P., Extremera, N. Y Ramos, N. (2004). Validity and reliability of the Spanish modified version of the Trait

Meta-Mood Scale. *Psychological Reports*, 94, 751-755. https://doi.org/10.2466/pr0.94.3.751-755

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología [Factor analysis as a technique in psychological research]. *Papeles del Psicólogo, 31*(1), 18–33.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica, 21*(3), 301–323.

Frías-Navarro, D., & Pascual Soler, M. (2012). Exploratory factor Analysis (EFA) in consumer behavior and marketing research. *Suma Psicológica, 19*(1), 47-58.

Grusec, J. E. (2011). Socialization processes in the family: Social and emotional development. *Annual review of psychology, 62*, 243-269. https://doi.org/10.1146/annurev.psych.121208.131650

Hernández, M. M., Eisenberg, N., Valiente, C., VanSchyndel, S. K., Spinrad, T. L., Silva, K. M., ... & Southworth, J. (2016). Emotional expression in school context, social relationships, and academic adjustment in kindergarten. *Emotion, 16*(4), 553. https://doi.org/.1037/emo0000147

Kar, N., & Kar, B. (2017). Social cognition and individual effectiveness in interpersonal scenarios: A conceptual review. *Journal of Mental Health and Human Behavior, 22*, 27-34. https://doi.org/10.4103/0971-8990.210705

Kvavilashvili, L. & Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: A critical and historical review. History & Philosophy of Psychology, *6,* 59–80.

Laghi, F., Lonigro, A., Pallini, S., & Baiocco, R. (2018). Emotion regulation and empathy: Which relation with social conduct? *The Journal of genetic psychology, 179*(2), 62-70. https://doi.org/10.1080/00221325.2018.1424705

Lewis, M.; Sullivan, M.W.; Stanger, C. & Weiss, M. (1989): Self-development and self-conscious emotions. *Child Development, 60*, 146-156. https://doi.org/10.2307/1131080

Lievens, F. R. O., & Chan, D. (2017). Practical Intelligence, Emotional Intelligence, and Social Intelligence. In *Handbook of Employee Selection* (pp. 342-364). New York: Routledge. https://doi.org/10.4324/9781315690193-15

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426-441. https://doi.org/10.1108/00483480810877598

Lorenzo-Seva, U. y Ferrando, P. (2007). *FACTOR: A computer program to fit the exploratory factor analysis model*. University Rovira y Virgili. https://doi.org/10.3758/BF03192753

Lorenzo-Seva, U. y Ferrando, P. (2011). Manual of the program FACTOR v. 8.02. https://psico.fcep.urv.cat/utilitats/factor/documentation/Manual_de_Factor_Esp.pdf

MacCann, C. y Roberts, R. D. (2008). New paradigms for assesing emotional intelligence: theory and data. *Emotion, 8*(4), 540–551. https://doi.org/10.1037/a0012746

MacCann, C., Wang, L., Matthews, G. y Roberts, R. (2010). Emotional intelligence and the eye of the beholeder: comparing self- and parent-rated situational judgments in adolescents. *Journal of Research in Personality, 44*, 673–676. https://doi.org/10.1016/j.jrp.2010.08.009

Matthews, G., Zeidner, M., & Roberts, R. D. (2017). Emotional intelligence, health, and stress. *The handbook of stress and health: A guide to research and practice*, 312. https://doi.org/10.1002/9781118993811.ch18

Mayer, J. D., Salovey, P. y Caruso, D. R. (2002a). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT): item booklet*. Multi-Health Systems. https://doi.org/10.1037/t05047-000

Mayer, J. D., Salovey, P. y Caruso, D. R. (2002b). *The Mayer-Salovey-Caruso*

Emotion lntelligence Test SCE T User's manual. Multi-Health Systems.

Monnier, M. (2015). Difficulties in defining social-emotional intelligence, competences and skills-a theoretical analysis and structural suggestion. *International journal for research in vocational education and training, 2*(1), 59-84. https://doi.org/10.13152/IJRVET.2.1.4

O'Connor, P. J., Hill, A., Kaya, M., & Martin, B. (2019). The Measurement of Emotional Intelligence: A Critical Review of the Literature and Recommendations for Researchers and Practitioners. *Frontiers in Psychology, 10*, 1116. https://doi.org/10.3389/fpsyg.2019.01116

Rodrigo, D. (2017). *Evaluación de la inteligencia emocional capacidad en adolescentes españoles: adaptación y validación del Situational Test of Emotion Management (STEM)*. (Tesis doctoral) Universidad Nacional de Educación a Distancia (UNED), España. https://doi.org/10.21134/rpcna.2018.05.2.6

Rodríguez-Ruiz, D. (2017). *Evaluación de la inteligencia emocional capacidad en adolescentes españoles: adaptación y validación del Situational Test of Emotion Management (STEM)* [Tesis de doctorado, Universidad Nacional de Educación a Distancia (España)] http://e-spacio.uned.es/fez/view/tesisuned:Educacion-Drodrigo

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: origins and outcomes. *Journal of personality and social psychology, 78*(1), 122. https://doi.org/10.1037/0022-3514.78.1.122

Saarni, C. (1997): Emotional Competence and Self-Regulation in Childhood, 35-64. In Salovey, P. & Sluyter, D. (eds.). *Emotional Development and Emotional Inteligance. Educational Implication.* BasicBooks.

Saarni, C. (2000). Emotional competence: A developmental perspective. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 68-91). Jossey-Bass.

Saarni, C. (2008). The interface of emotional development with social context. In M. Lewis, J. Haviland-Jones, & L. Feldman Barrett (Eds.), *The handbook of emotions* (3rd ed., pp. 332–347). Guilford Press.

Sala-Roca, J., Filella, G., Oriol, X., Ros, A., Secanilla, E., Rodríguez, M., Soldevila, A. (2016). *Test Situacional de Desarrollo de Competencias Socioemocionales (DCSE-J)*. Versión revisada por Josefina Sala, Anna Soldevila, Sara Rodríguez Pérez, Aida Urrea-Monclús y Nair Zárate en 2020. Universitat Autònoma de Barcelona. https://ddd.uab.cat/record/212951

Salovey, P., Mayer, J.D., Goldman, S.L., Turvey, C. y Palfai, T.P. (1995). Emotional attention, clarity and repair: exploring emotional intelligence using the trait meta-mood scale. En J. Pennebaker (ed.): *Emotion, disclousure and health*. Washington, D.C.: American Psychological Association. https://doi.org/10.1037/10182-006

Sánchez-Nunez, M., Fernández-Berrocal, P., Montañés, J., & Latorre, J. M. (2008). Does emotional intelligence depend on gender? The socialization of emotional competencies in men and women and its implications.

Sánchez-Teruel, D. y Robles-Bello, M.A. (2018). Instrumentos de evaluación en inteligencia emocional: una revisión sistemática cuantitativa. *Perspectiva educacional. Formación de profesores, 57*(2), 27-50. https://doi.org/10.4151/07189729-Vol.57-Iss.2-Art.712

Sharma, S.; Gangopadhyay, M.; Austin, E. y Manas, K.M. (2013). *Development and Validation of a Situational Judgment Test of Emotional Intelligence. International Journal of Selection and Assessment, 21*(1), 57-73. https://doi.org/10.1111/ijsa.12017

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual review of clinical psychology, 11*, 379-405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and Reliability of Situational Judgement Test Scores A New Approach Based on Cognitive Diagnosis Models. *Organizational Research Methods, 19*(3), 506-532. https://doi.org/10.1177/1094428116630065

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in cognitive sciences, 9*(2), 69-74. https://doi.org/10.1016/j.tics.2004.12.005

Authors / Autores
Rodríguez-Pérez, Sara (rodriguezpsara@uniovi.es) 0000-0001-6964-9156
Assistant Professor of the Department of Educational Sciences of the University of Oviedo. She is a teacher, pedagogue and sexologist. She coordinated and developed the Asexora project (Sexological Counseling for young people) in collaboration with the Youth Council of the Principado de Asturias. She has participated in research projects on intervention with families in educational spaces, influence of gender in academic decisions, well-being of care leavers, emotional education and Dual Professional Training. She is currently part of the IARS group (Childhood and Adolescence at Social Risk). Instagram: medrando_con_buenos_tratos
Sala-Roca, Josefina (Fina.Sala@uab.cat) 0000-0001-9829-5131
Professor of the Department of Theory of Education and Social Pedagogy of the Autonomous University of Barcelona. She coordinates the Research Group on Childhood and Adolescence at Social Risk (IARS). The IARS group has developed various studies on the development of socio-emotional and employability competencies of children and young people. Also, in collaboration with FEPA, FEDAIA and DGAIA among others, it has developed studies on the factors that affect the well-being and positive development of young people in care, as well as the difficulties and supports in their transition to independent life. The research carried out has been published in various articles and reports. Twitter: @IARS_uab
Doval-Diéguez, Eduardo (Eduardo.Doval@uab.cat) 0000-0001-8416-160X
Professor of the Department of Psychobiology and Methodology of Health Sciences at the Autonomous University of Barcelona. His main research focuses on the identification and study of aberrant response patterns in educational assessment tests. He is a member of the GIES group (Research group in stress and health).
Urrea-Monclús, Aida (aida.urrea@udl.cat) 0000-0002-8661-9191
Acting Associate Professor of the Department of Pedagogy of the University of Lleida. Her research and teaching focuses on the promotion of children's rights both in the child protection system and in the educational system. Currently, she participates in research to promote the participation of young people in the protection system, especially in their transition to independent life and in family reunification processes. She is part of the IARS Group, the GRISIJ Group and the Education and Adolescence Chair. Twitter: @aida_aum

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa

E-Journal of Educational Research, Assessment and Evaluation